Boston College
Lynch School of Education and Human Development

Department of Measurement, Evaluation, Statistics, and Assessment

TOWARDS HOLISTIC EVALUATION OF EDUCATION SYSTEMS: USING TIMSS 2023 CONTEXT DATA TO CLASSIFY SCHOOLS BY SCHOOL CLIMATE HEALTH

Dissertation

by

CHARLOTTE ELIZABETH ARMBRUST ALDRICH

submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

August 2024

Boston College, Chestnut Hill, MA

ABSTRACT

Children internationally are entitled to quality education. International education initiatives monitor education system quality through complex evaluations, historically relying on academic benchmarks operationalized by robust comparative achievement data. However, quality in schooling is evolving to comprise development beyond academic abilities – it should support emotional, social, and psychological development. Valid systems-level evaluations of these features internationally require well-defined benchmarks for school conditions suitably supportive of this development. Emerging international initiatives, such as UNESCO's Happy Schools Framework, define frameworks for non-academic facets but have not been empirically tested. This study defines *school climate health* as the intersection of the Happy Schools Framework, existing literature on school climate and wellbeing, and the Trends in International Mathematics and Science (TIMSS) Context Questionnaire.

This research aims to provide a first step towards defining benchmarks by exploring an international dataset to define existing patterns of interrelated school context variables. This study is responsive to empirical literature and relevant theoretical frameworks for evaluating social systems (systems evaluation, ecological systems theories). An exploratory multilevel latent class analysis (MLCA) of 22 variables is conducted for the 58 participating countries to define four school clusters and three country classes defining the composition and distribution of school climate health internationally. Combining response variables from students, teachers, principals, and parents is a novel application. Characteristics of each school cluster and country class are described. Secondary analyses investigate possible confoundedness of school demographics and possible relationships between school-level average achievement.

ACKNOWLEDGMENTS

To say I acknowledge all those I owe thanks and gratitude to is to undersell the support I have been lucky to have. I have been fortunate to find encouragement, enlightenment, and friendship throughout my doctoral experience. It has sometimes been a struggle, but it has also been a privilege to spend this time pursuing a single idea from fruition to finale.

To my dissertation chair, Matthias von Davier, I am thankful for your confidence that I could take this giant project to the end – thank you for reviewing my writing and guiding me as I learned. To Ina Mullis, I am grateful for bringing me into TIMSS and PIRLS in the first place and for your encouragement and mentorship. Emily Gates, I believed I could think big and strike fast learning from you – I'll never think of 'value' the same way. Belle Liang, thank you for our conversations about school community and relationships that are at the heart of this dissertation.

Working full-time during my doctoral studies has been challenging, but it has connected me with a diversely talented group that has buoyed my spirits countless times. I am so thankful to the TIMSS & PIRLS International Study Center – whether traveling the world or connecting in the office. Over seven years, I cherish these relationships and your professionalism, diligence, and creativity.

I am fortunate to have loved ones who remind me that there is more than the grind. My parents were always convinced I could be a doctor if I just kept going. I'm thankful for my brother, also pursuing a doctorate, for our trips to visit each other to commiserate and bake sugar cookies. My husband, Jack, is truly a partner for the ages – we've hit so many milestones during this time and there is only more good to come. My in-laws, the Swanson Extended Universe, make the Commonwealth feel like home. To my friends and classmates, thank you for snapping me out of it when needed and pushing me back into the thick of it as necessary.

Thank you, everyone.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Children deserve to be well and thrive as a general principle of a humane world, both in their home lives and at school. However, as part of policy implementation toward addressing a public concern, it is not appropriate to set accountability targets to maximize all measures at once - to have the highest of all the "good" measures. Missing in all levels of the literature is a theorization of a satisfactory state, an idea of what is an adequate level of thriving. Such a theorization recognizes that not all humans can grow and thrive at the same maximum level all the time. In educational contexts, any individual could name and isolate some qualities of an ideal state they would like to see in a school where they are involved. But as the number of stakeholders and their interactions increase so does the complication of identifying these qualities. With this increase, though, all the greater is the potential for diversity of ideas and values that expand the possibilities of what schooling could be or accomplish.

International comparative education is grounded in such a collaborative ideal, that opening the field up to many best practices is worth the complication of finding agreement. A longstanding relationship exists between different levels of national education systems and International Large-Scale Assessments (ILSAs). These projects such as the Trends in International Mathematics and Science Study (TIMSS) conducted by the TIMSS & PIRLS International Study Center at Boston College for the International Association for the Evaluation of Educational Achievement (IEA) have been a crucial data source to answer the questions posed about the state of education internationally. As school quality accountability programs have experienced a shift towards inclusion of non-academic school outcomes like student mental health and overall school climate, such programs are called on to track more than academic achievement as well.

The shift from academic achievement towards non-academic outcomes needs to be examined in light of the fundamental idea of comparability driving international studies such as TIMSS. Regarding academic content, the assumption is that whatever cultural differences there may be in communicating that content are manageable such that a comparison of comprehension of mathematics and science principles is possible across countries. Turning to non-academic qualities, the focus of that measurement shifts from cognitive tasks to personal attributes and behaviors sensitive to cultural differences. How one feels about these or acts them out is intrinsically linked to cultural norms and expectations. Such inherent difference brings into question whether a basis agreement on the types and value of these attributes can even be established to make comparisons possible. Further complicating this shift is that it is untested how to meaningfully evaluate these systems-level aggregations of interpersonal attributes.

The present study confronts these problems by combining relevant scales relating to perceptions of the school environment using methodologies that can capture complex structures of interdependencies. Such a collection and combination of scales could support the characterization of *school climate health*: the extent of a productive social environment that sustains student learning and well-being alike. The resulting profiles of school climate health simultaneously capture the state of key environmental factors supportive of student wellbeing and learning such as student relationships, teacher performance, and school safety. 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 those involved. School climate health is a holistic understanding of school well-being supported by the literature and is operationalized with a selection of TIMSS 2023 Grade 4 Context Questionnaire scales.

The TIMSS Context Questionnaires have great potential as a base of inquiry in examining some of these non-academic qualities. There is untapped opportunity therein to examine whether and how attributes relevant for describing school experience that drive student wellbeing can be compared internationally in terms of non-cognitive outcomes. Whereas stakeholders using TIMSS achievement data can rely on an overarching measure of mathematics and science ability per country, a theory or at least a working hypothesis is needed to meaningfully combine the data from multiple scales available in the context questionnaires. While the achievement items are designed to form a scale representing a single mathematics or science domain, context questionnaire items either stand on their own or are combined into short scales and reported as a multitude of background variables.

TIMSS could report a national representation of the multidimensional circumstances of school life. Fundamental to this dissertation is that appropriately aggregating these data from different levels of schooling context could allow for a comprehensive understanding of the complexities of school environments. In that case, it is possible that TIMSS can support large-scale evaluations of the quality of schooling at the national level. Deriving patterns of well-being and school climate based on scales available at the different levels TIMSS collects context data is expected to provide support for answering 'big-picture' questions about the quality of national education as measured by the school environments sustained in the country.

Purpose and Research Questions

The purpose of this research is to derive patterns from a selection of scales administered to students, parents, teachers, and principals as part of the TIMSS 2023 Grade 4 Context Questionnaires that are related to school climate health. This exploratory study endeavors to understand the defining similarities of school social systems internationally. Assuming that the

scales that are commonly used to evaluate wellbeing and school climate health can meaningfully separate schools with respect to these constructs, the study aims to understand how the different groups of schools are different and what that might mean for the quality of education they produce. The research is responsive to fields that demand more integration across stakeholders and leverages the theorization that school climate and school well-being are dynamic social systems involving stakeholders with different values and roles to act out.

This dissertation sources its structure from a conceptualization of school climate found in the Happy Schools Framework developed by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) and several Asia-Pacific countries. The Happy Schools project endeavors to implement sustainable school systems that support student academic success and well-being, grounded in a whole-school approach to "prioritize happiness and well-being" for "better quality" schooling (UNESCO Bangkok, 2020). The original three-part framework comprises several facets that will "promote happiness" by way of well-being and holistically supporting student development, pointing to SDG 4.1 and SDG 4.7 as justification for the endeavor. While the Happy Schools framework has not been empirically tested, existing empirical measures with similar grounding are either not specific to school environments or not designed to be internationally comparable

The extent that subdomains from the Happy Schools framework are related, yet not empirically integrated, is where the present research fits into the state of the field. This research used the selected scales of the Grade 4 Context Questionnaires to distill existing frameworks for non-academic school environment constructs. The scores on these scales serve as the indicator variables for identifying groups of schools among those participating in TIMSS 2023. Although,

since the TIMSS 2023 Contextual Questionnaire has already been finalized, constructs covered by this study must have adequate alignment with the existing scales being administered.

Research Questions

1. To what extent do TIMSS 2023 Grade 4 Context Questionnaire scales map to internationally established constructs of school climate and student well-being?
2. What patterns can be derived internationally from the selected response data?
3. What qualities do schools share within school climate health groups, what distinguishes the members of one group from others, and how does group membership relate to mathematics achievement?

Background Literature

Several international initiatives continue to ask for the prioritization of non-cognitive characteristics such as well-being and school climate in large-scale evaluations of education quality in schools and schooling. For instance, the Sustainable Development Goals (United Nations, 2015) call not only for education systems to produce high-achieving students but also for the "promotion of a culture of peace and non-violence, global citizenship and appreciation of cultural diversity and culture's contribution to sustainable development" (SDG 4.7.1; United Nations, 2015). More broadly, there is support for a greater focus on non-academic measures of academic success as awareness surrounding the mental health and well-being of students grows. Pedagogically, the influence of Whole Child Education has raised awareness of the value of socio-emotional development and well-being through education (Shirley, 2020; Wortham et al., 2020). In educational assessment and evaluation, there are calls for more indicators of these non-academic outcomes to support continuous change and identify localities for intervention (Wei, 2015). The conditions of a school's environment act upon students and impact academic and

non-academic outcomes alike. While measures of student emotional development have been created in response to an increased interest in non-academic indicators of school quality (Muller et al., 2020), student-level measures are limited in scope as measures of overall education quality. Looking at the student level alone leaves any effects of the other actors unexplored. Further, claims based only on the student attributes or behaviors can have the unintended implication that students are solely responsible for the outcomes of their schooling despite the numerous influences acting on them throughout.

Non-academic outcomes like 'well-being' (e.g. Diener et al., 2018) and 'school climate' are understood more as a system of social actors that work to either sustain or inhibit a positive school climate (e.g. Thapa et al., 2013; Allen et al., 2021). One route to measure how these actors are playing out their roles is to capture their perceptions of "the social interactions and relationships, sense of physical and social-emotional safety, and the values and beliefs held by students, teachers, administrators, and/or staff within a school" (Rudasill et al., 2018, p.46). This implies that the quality of a school climate may be assessed by how the individuals describe it to be, rather than based on descriptive data like demographics and personal characteristics

Conceptual Framework – Supporting Evaluation of Education Systems

Evaluation is a key feature of the lifespan of large-scale government initiatives to sustain good practices and accountability, establish a culture of learning, and illuminate previously unnoticed public failings (Chelimsky, 2006). A key argument of this research is that evaluations of children's well-being internationally require the development of criteria that are internationally valued. Values define what is "intrinsically good, desirable, important, and of general worth" within the context of the evaluand. (Davidson, 2014). One can use evaluative rubrics to qualitatively describe what "evidence should look like at different levels of

performance" for the valued outcome, and then draw conclusions based on the evidence (Davidson, 2013). In defining what is valued within the evaluation context, a process referred to as valuing, it is essential to incorporate numerous perspectives from stakeholders that operate or benefit from the entity being evaluated (Alkin, Vo, & Christie, 2012).

Making an ultimate evaluative judgment based on these criteria is a defining requirement of the field of evaluation. "The analytic approach to valuing requires distinguishing and selecting the criteria, sometimes thought of as dimensions, on which the thing, the evaluand, is to be judged" based on the evidence gathered towards a conclusion (Julnes, 2012). Evaluating the performance of an evaluand is a reasoned argument combining several indicators to respond to the needs of the question of quality or value being asked. However, evaluation can do more than assign worth. Evaluation can highlight strengths in an underperforming entity and weaknesses in a successful one towards sustaining improvements (Davidson, 2014). Conducting an evaluation is an integrative project that arrives at conclusions beyond data reporting alone. Not only is there a need for a framework, but to conduct evaluation there must also be definitions of varying levels of suitability.

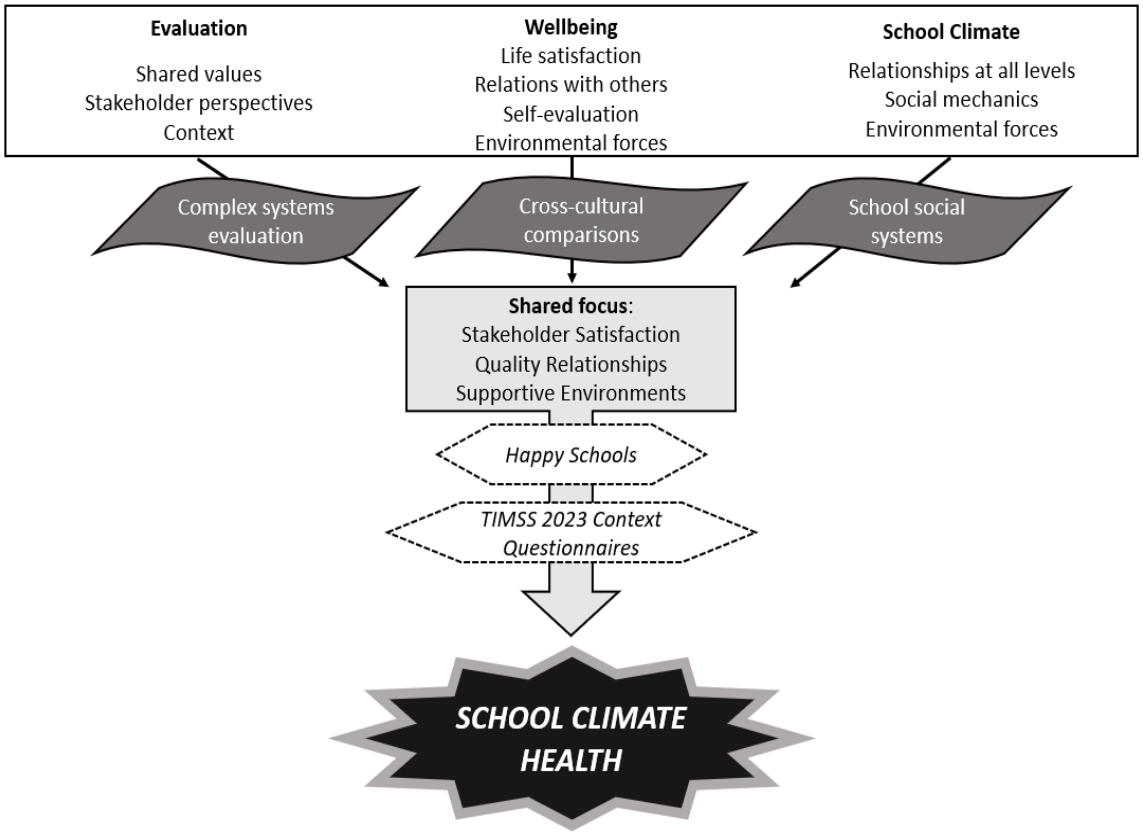
School quality accountability programs should include more information than just academic achievement. International bodies are setting goals about being good citizens of the 21st century. What approach, with what data, can all this be answered? The commonality of what school quality entails based on varying conceptions of environmental characteristics is unclear, as are the appropriate methods to summarize them. Evaluations of qualities such as student well-being and school climate require measures or rubrics comprising input from multiple points of view, such as students, their parents, teachers, and principals, to support a comprehensive understanding. To respond to systems-level evaluative inquiries, the methods of combining

responses from students, principals, and parents must be reimagined to comprehensively capture the state of school climate and wellbeing internationally.

There is interest in and support for multiple perspectives that represent the facets of a school's social system. However, those existing measures are not written with an international sample in mind and do not aggregate perspectives from multiple stakeholders. School climate health represents a synthesis of numerous literature bases, filtered in two stages through existing frameworks and measures. This multistage conceptualization is represented in Figure 1.1. The literature review of this study revealed a shared focus in evaluation, wellbeing, and school climate theories on levels of satisfaction, qualities of interpersonal connections, and considering contextual or environmental forces.

Figure 1.1

Visualization of School Climate Health Definition



For this research, special consideration in these fields for evaluating complexity, measuring wellbeing across countries, and understanding schools as social systems brings these shared interests into sharper focus. The Happy Schools Framework was developed to be responsive to many of these considerations, thus leveraging those criteria in response to the literature synthesis is the first step in operationalizing the school climate health. Finally, relevant criteria from Happy Schools are matched to related items and scales from the TIMSS 2023 Context Questionnaires. While empirical measures for multidimensional definitions of wellbeing and school climate have been adapted to numerous contexts, the TIMSS 2023 Context Questionnaires are designed specifically to measure conditions of schooling based on international agreement.

This study hypothesizes that not only can school climate be defined by student perceptions of the other actors but that the perceptions of multiple actors, operating at multiple levels in the school, can be combined to measure school climate health. Evaluations require benchmarks, some criteria against which to judge the characteristics of observed phenomena. In general, valid evaluations require criteria meaningful to the stakeholders to compare against evidence gathered in the course of the evaluation. Beyond descriptive research or assessment, to evaluate is to consider evidence alongside valued criteria (Davidson, 2013). An initial problem to be discussed is whether such an internationally agreed upon set of values exists, or can be created, that allows valid evaluations of student wellbeing and school climate across countries. TIMSS is based on the notion that international agreement is possible in both academic and contextual features of education worldwide. The content of the TIMSS Context Questionnaires represents a collection of important indicators beneficial to school environments as agreed on by the countries participating in TIMSS.

An additional problem is that discrete questionnaire scales do not necessarily directly respond to the holistic system-level goals set by organizations like the United Nations. Complexity is inherent to the large-scale undertaking of educating the world's children. Rather than reducing the complexity to a single measure, this dissertation takes on a somewhat maximalist approach to integrating a variety of viewpoints and factors to describe the state of education. All schools are a social network that operates only if all the person-level mechanisms are supported - a complex social system complex knowable by the roles and mechanisms that are executed within it (Bronfenbrenner, 1989; 2000). Schools with healthy school climates could simultaneously support academic achievement and personal and social development. Furthermore, if the mechanisms to support an environment fostering personal or social development can be isolated, then it is possible new pathways to improved academic attainment would also be unlocked.

Data and Research Methodology

The study has four major phases in responding to the three research questions:

A. Establish a framework of "healthy school climate" that can be measured by the TIMSS 2023 Context Questionnaires
B. Find clusters of schools internationally based on multilevel latent class analysis (Vermunt, 2003) of the metrics identified from previous.
C. Define clusters based on school qualities shared or distinguishing from others. Where are schools of e.g. Group A mostly found?
D. Explore the relationship between academic achievement and different categorizations of healthy school climate

TIMSS Context Questionnaire data is rich with information on the environments for learning that students operate within internationally and is the basis for all analyses in the study. Students are asked about the conditions of their schooling, how they perceive their relationships with their peers and teachers, and how they feel about being at school. Teachers are asked about their training and experience with professional development, but also their satisfaction with teaching as a profession, their feelings about the school, and the extent they feel supported in the task of teaching. Principals report on demographic information about their schools but also on the state of discipline and relationships in their school. Parents of grade 4 students are asked about early learning milestones for their children, but also the extent that they are satisfied with their school.

TIMSS is nationally representative and supportive of international comparison of education systems. There is an opportunity to explore what relationships might exist using these nationally representative samples. Context questionnaire scales from various stakeholders are reported through their relationships with the sampled students. Participating countries and benchmarking entities routinely use TIMSS results as part of national quality accountability. The purpose of the TIMSS context questionnaires is to gather information internationally on factors of educational environments that impact mathematics and science achievement as measured by TIMSS. The Context Questionnaire framework is established based on what is "generally considered to be important aspects of education systems and beneficial for student learning" in addition to "other topics that are important to TIMSS participating countries and education researchers but have not been shown to be related to achievement" (Hooper et al., 2018, p.4). Even though there will not be coverage of everything that literature or national entities find important, existing TIMSS data could do more than it is doing and could meaningfully describe

what nations actually agree on in terms of non-academic outcomes that are important in educating children worldwide.

Significance

The non-academic ambitions of international education initiatives lack an empirically supported framework and functional benchmarks. Taking from principles of evaluation, the framework proposed for school climate health attempts to work within what is available through the TIMSS 2023 Context Questionnaires to engage stakeholder perspectives, gather evidence from multiple stakeholder groups, and meaningfully integrate the data into profiles that describe the state of school environments internationally. The study is exploratory but hypothesizes that similarities on these chosen variables are similar enough internationally to describe what kinds of school environments exist. It is important to look at the environmental level of the school to embrace the socio-ecological conceptualizations of school climate and student wellbeing that are in the literature, but also to promote sustainability inherent to the SDGs and initiatives looking for "long term impacts".

An unknown this research faces is what is reasonable to compare internationally and what levels of different school quality metrics exist in schools. Without this information, it is hard to set benchmarks for progress or evaluate the distance to reaching a "satisfactory" school environment, or understand multiple "satisfactory" states that exist in the way schools are operating internationally. This study is an exploratory first step towards a possible analytic pathway to respond to the demand in numerous fields for greater integration of multiple perspectives into the modeling of school climate.

With the continued presence of the COVID-19 pandemic, institutions internationally are pressingly concerned with regaining achievement gaps as well as the condition of students'

mental states (Shirley, 2020). There is considerable urgency to make up for lost time– might considering student well-being be a distraction from what schools should teach children? This research offers a possible path to meaningfully embracing systems thinking in evaluating the non-academic qualities of schools internationally. What both the literature on non-academic outcomes and large-scale educational evaluations share is a call to embrace systems thinking or complex systems theory as a lens to capture the complexity of education at large. If we apply that lens to the longstanding relationship ILSAs have in national accountability systems, what can TIMSS in particular provide to support systems-level evaluations of school quality? Some frameworks exist that suggest the experiences of multiple stakeholders must be considered to create sustainable school environments to support student well-being. Further, measuring student well-being in one instance is not enough to evaluate the long-term sustainability of schooling outcomes.

The outcome of this research supports evaluations of education quality by proposing both a framework and a possible emergent rubric for evaluating school climate health. The present research broadens conceptualizations of school well-being in response to demand in the literature for well-being, school climate, and evaluation to approach big problems with systemic methods. The results will clarify what defines school environments internationally beyond descriptive reporting to understand the state of school environments that international initiatives aim to correct. By exploring the relationship between school categorization and academic achievement as measured by TIMSS, the research challenges the assumption that achievement is indicative of school quality alone.

This work also supports the ongoing continuous improvement of the TIMSS program of study. While the questionnaires have been finalized for the TIMSS 2023 cycle, the present

research focuses on how national-level reporting on these existing scales can be construed differently than the separate indicator approach that is predominantly used in reporting and secondary analyses. Traditionally, TIMSS focused on results on the achievement scales and limited explanatory explorations to marginal tables that showed how select context scales relate to achievement measures. Instead of separating the scales by type of respondent, i.e. separate scales for teachers, parents, and students, this study is aimed at supporting holistic evaluation of school climate by combining indicators thematically across these levels.

Chapter Summary

To the extent that TIMSS Context Questionnaires target what is important at the systems level of school quality, data from various stakeholder respondents could be integrated to describe the baseline that would be necessary for evaluations of school quality. As a first, exploratory estimation of what is possible, this research used data from multiple stakeholders to comprehensively describe the present state of school environments internationally. A selection of TIMSS 2023 Grade 4 Context Questionnaires are selected based on school climate and student well-being literature to define school climate health. The remainder of this dissertation will follow through a review of that literature, a description of the methods, followed by results, and a discussion including limitations.

CHAPTER 2: LITERATURE REVIEW

This chapter argues that the qualities of school environments are essential and inherent to capturing student wellbeing. School climate health relies on a unifying hypothesis of what non-cognitive attitudes students are entitled to having their school support. Beyond skill development, there is also an emphasis on the psychological health of students, with terms like mental health, wellbeing, and life satisfaction being of concern. From a diverse literature base, the need to understand the systemic interactions and influences acted out by different kinds of stakeholders is clear.

Sustaining student wellbeing demands a comprehensive definition of the condition of school functioning as a productive social system. In this study school climate health can be described as a multidimensional profile of the qualities of relationships, degree of satisfaction, and extent of supportive school environments that are internationally valued. This research proposes an analytic framework and an emergent methodology to develop criteria for that construct. The progression of this chapter will provide an overview of recent global initiatives to respond to shifting understanding of school quality and summarize the predominant theorization for wellbeing and definitions of school climate. The need for benchmarks aligned with validated, systems-informed frameworks that capture the interrelatedness inherent in evaluating large-scale social initiatives is argued. Expanding the boundaries of wellbeing accordingly lends itself to a discussion of school climate and the role of systems theory in conceptualizing the dynamic mechanics of school operations captured by school climate health.

International Initiatives

The wellbeing of children is a longstanding matter of public concern that evolved from a history of orphans being neglected and children being put to work in harsh conditions that

deteriorated their health (Doek, 2014). Beyond local government, international bodies have been involved in monitoring childhood conditions globally, first with the League of Nations and then its successor, the United Nations (UN). It is widely agreed upon that the welfare of children is an international priority and that they have an inherent right to participate in quality education (UN General Assembly, 1990). The UN Convention on the Rights of the Child (CRC) was and continues to be the human rights document ratified by most countries ever. Commitment to accountability for conditions of childhood has sustained the growth of international monitoring of academic attainment and bolstered the field of international comparative education. However, to have evaluative comparability in monitoring student wellbeing worldwide, a working framework that is simultaneously comprehensive and responsive to international differences in values has yet to be satisfactorily developed.

International initiatives from organizations including the UN have long included educational ambitions and targets for member countries to pursue in their national policies. Within a decade of the CRC, the UN established the Millennium Development Goals (MDGs) as the key goals for nations to jointly pursue shared values of international cooperation (UN General Assembly, 2000). Most recently, the UN's Sustainable Development Goals (SDGs) build on the achievements of the MDGs, going beyond supporting shared values to addressing a shared responsibility for the future of humanity (UN General Assembly, 2015). Comparing the framing of these two initiatives, a shift in discourse and theorization throughout the early 2000's is evident. While the MDGs all relate to a global ambition to end poverty, hunger, and disease, each goal is presented as a distinct building block towards that overall objective. The phrasing of the SDGs is heightened by an acknowledgment of the complexity of achieving change and that urgency is required in advancing them. The goals of SDGs reference the shortcomings of the

MDGs and pursue "the bold and transformative steps which are urgently needed to shift the world onto a sustainable and resilient path" (UN General Assembly, 2015). The components of the SDGs are understood to be interrelated, influence each other, and must also sustainably transform the world for the future.

Comparing the goals for education across the two initiatives is similarly illuminating. For the MDGs, the goal was to ensure that all children have access to primary school. For the SDGs, though, there is much more specification of what that education should look like across the seven education-specific goals, evident in the seventh goal (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 (United Nations General Assembly, A/RES/70/1)

Target 4.7 is concerned with more than literacy and numeracy, going beyond content-specific measures of school quality to state that there is a certain outlook and ethos students should learn as part of their schooling. The UN tracks progress on this goal based on how integrated Global Citizenship Education (GCED) and Education for Sustainable Development (ESD) are in the national curriculum, teacher preparation, and student assessment (UN Statistics Division, 2022) There are eight shared principles that the UN asks member countries to report in terms of level of integration into their education policy documents: cultural diversity and tolerance, gender equality, human rights, peace and non-violence, climate change, environmental sustainability, human survival and wellbeing, and sustainable consumptions and production. Each of these

themes is measured at various levels of implementation to see if "the basic infrastructure exists that would allow countries to deliver ESD and GCED to learners, to ensure their populations have adequate information on sustainable development and lifestyles in harmony with nature" (UN Statistics Division, 2022, p.3). This is the extent of the official guidance for countries to inform national evaluations of progress in implementing SDG 4.7.

Precedence exists for international large-scale assessment (ILSA) data to inform progress indicators on these global initiations. IEA studies, specifically TIMSS, have been cited as sources for tracking other indicators of SDG 4 (UNESCO, 2020; UNESCO 2022). UNESCO and IEA publications not only support the use of TIMSS data to track academic progress but also to monitor school experiential indicators such as SDG 4.a.2 which relates to the frequency of student bullying (UNESCO, 2021; IEA, 2022). This target in particular was defined after the initial publication of the SDGs to enhance the measurement of progress towards SDG 4.a, included along with 4.a.2, 4.1, and 4.7 in Table 2.1. (UNESCO, 2019) Thus, there is support and recognition of TIMSS data as part of international monitoring of these large-scale educational objectives. TIMSS data has already been used in UNESCO national progress reports for SDG indicators such as 4.1.1 (UNESCO 2018). However, all these instances rely on the extent that the comparisons being made are based on targets "whose definition does not depend on country or political agendas, but rather can be guides for all countries, regardless of the widely varying contexts" (Hadstedt, 2020; Pineda & Sandoval-Hernandez, 2023). In other words, the supportability of TIMSS data for one internationally comparative application is only reassuring to the extent that a comparable basis can be found for another. In the case of this research, leveraging TIMSS data appropriately requires an understanding of what framework can form the basis for evaluating more complex objectives.

Table 2.1

Select SDGs and Indicators

Objective	Overview	Indicator
4a	Build and upgrade education facilities that are child, disability, and gender sensitive and provide safe, non-violent, inclusive, and effective learning environments for all.	4.a.1 Proportion of schools with access to: (a) electricity; (b) the Internet for pedagogical purposes; (c) computers for pedagogical purposes; (d) adapted infrastructure and materials for students with disabilities; (e) basic drinking water; (f) single-sex basic sanitation facilities; and (g) basic handwashing facilities (as per the WASH indicator definitions)
		4.a.2 Percentage of students experiencing bullying in the last 12 months
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

Evaluating the SDGs – Complexity and Systems

EvalPartners, an international consortium of professional evaluators affiliated with the UN, was established to advocate for responsive monitoring processes as part of international programs toward achieving the SDGs. Among their many activities is a series of briefings on possible routes for evaluating progress toward the SDGs. In part, the briefings are introductory with information such as what distinguishes evaluation from assessment or research (Schwandt et al., 2016a). Not only is evaluation necessary to promote learning as nations work towards the SDG targets, but those evaluations should embrace the complexity inherent to the interconnectedness of the targets (Lucks, Etta, & Miranda, 2019). For example, achieving the education-specific goal of 4.7 mentioned previously is also an indicator of SDGs 12 (Ensure sustainable consumption and production patterns) and 13 (Take urgent action to combat climate change and its impacts) (UN, 2015). Evaluation that integrates methodologies of systems thinking and complexity science is designed to manage these context-dependent entanglements of social issues, often referred to as "wicked problems" (Longrenn & van Poeck, 2021). Evaluation can also be a process to develop a better understanding of such phenomena, rather than a project with a fixed conclusion (Ofir et al., 2016). Engaging in systems evaluation supports multiple perspectives within the defined boundaries of the system in question (Schwandt et al., 2016b). Furthermore, understanding how the system works well enables evaluators and stakeholders to directly interrogate what sustainability of system functioning requires (Zazueta et al., 2021). Systems evaluation responds to the interconnectedness of modern problems while still maintaining the importance of context, evaluative thinking, and stakeholder participation.

Calls for systems evaluation are often tinged with urgency and respond to the need to address pressing problems requiring "transformational change" or "long-term impacts" (Ofir and

Rugg, 2021). Theories of complexity also influence systems theory to reject causal, linear theories of change for a holistic lens to understand the norms that govern the interrelatedness of systems comprising people (Gates et al., 2021). Systems evaluations do not have a requisite methodology but leave room for emergent states of being that reject deterministic theories of change (Ofir and Rugg, 2021). Developing these types of evaluation programs requires keen attention to boundary definitions to establish what entities have stronger linking relationships than those lying outside of the forces of the system (Gates, 2016). Through this practice, evaluators and stakeholders define ideal system states starting with conceptualizing its current state. Beyond what is necessary for traditionally conceived program evaluation, evaluations of systems like school social systems or national education apparatuses must also contend with the complexity and interrelatedness through holistic designs and reporting.

Happy Schools Project and Framework

The Happy Schools project from UNESCO endeavored initially to create sustainable school systems in Asia-Pacific countries that support student academic success and wellbeing, grounded in a whole-school approach to "prioritize happiness and wellbeing" for "better quality" schooling (UNESCO Bangkok, 2020). The three-part framework comprises several facets that will "promote happiness" by way of wellbeing and holistically supporting student development, pointing to SDG 4.1 and SDG 4.7 as justification for the endeavor. Created from the responses of a broad audience of students, teachers, principals, parents, and the general public, a qualitative, open-ended survey asked:

1. What can make school a happy place?
2. What can make school an unhappy place?
3. What makes teaching and learning in schools fun and enjoyable?
4. From your experience, what can be done to make sure that all students feel included in schools?

The original framework is split into three major constructs: people, process, and place. Supportive relationships are the basis of "people"; the quality of learning is the basis for all components in "process"; and the institutional practices and norms of the school are "place" (UNESCO Bangkok, 2016, p. xiv). Notably, the framework comprises the perceptions and attitudes of stakeholders beyond just students as participants in building a "happy" school. The framework has only been tested via a small-scale two-year operational pilot conducted from 2018-2020 in five schools each in Japan, Laos, and Thailand. While the Happy Schools framework has not been empirically tested, its potential for adaptation is untapped – there are empirical measures with similar grounding but that do not adequately account for the complexity of student wellbeing and international comparison.

Figure 2.1

Original Happy Schools Framework

Whether schools need to be "happy" and even what defines happiness could be subject to conflicting opinions internationally. Philosophically, the Happy Schools framework draws on definitions of happiness from Socrates and Confucius that are appropriate framing for its target region but may not be grounding that school systems share internationally. However, happiness can be understood to refer to wellbeing in this circumstance without losing meaning. Given the literature reviewed to this point, the components of the Happy Schools framework were selected in two stages for inclusion in this study. Namely, there is sufficient support for the systems-level conceptualization of student wellbeing (e.g. happiness) and school climate. Furthermore, there is agreement in the literature that the qualities of relationships among people in social systems, such as schools, define wellbeing and climate. Complexity is inherent, but manageable through a systems lens. These qualities can be reliably measured via self-reports of stakeholders' perceptions. However, these perceptions are inextricable from the role that a stakeholder plays in the system.

With the recent publication of additional Happy Schools reports, project authors emphasize that happy schools not only promote learning but also have inherent benefits beyond potential outcomes for academic achievement:

> Happy schools can be sites or spaces to support social cohesion, creating communities across differences. Happy schools, too, can foster a lifelong love of learning through joyful engagement, rather than pressuring academic performance over all else to the detriment of personal well-being, which will in turn undermine learners' engagement in learning. In short, happy schools can provide the safety, support, and positive social interactivity engagement that is needed for students to learn." (Mahfooz & Norrmen-Smith, 2022, p. 5)

Building on the original framework launched in Asia, the Happy Schools Initiative embraces a whole school approach and acknowledges the added urgency since its original publication given the emotional and academic declines since the COVID-19 pandemic. UNESCO has since launched an international version of the Happy Schools Framework (HSF) hoping to increase the profile of the importance of developing schools with positive climates (UNESCO, 2024). The international framework version has some adaptations from the original, maintaining the triptych of *People*, *Process*, and *Place*, and adding a category called "Principles". While the framework may group the components differently, the targets and criteria are resonant with the original and with the definitions in this study. The criteria are well-aligned between versions, with the new category representing a reformatting of the principles that everyone involved in schooling benefits from a shared, positive, productive learning environment.

The internationalization of the project has also prompted in-country studies to validate the definition of Happy Schools in specific country contexts. Recent publications from UNESCO re-emphasize the necessity for localization and invite adaptation to the Happy Schools framework to best suit a local environment. A pair of studies in Portugal investigated the extent that parents' and students' personal values for being 'happy' in school aligned with the criterion of HSF. The study found that the most valued happy school criteria among parents was their children's relationships with their peers, teacher professionalism, and student creativity (Gramaxo et al., 2023). When the same team of researchers investigated similar questions in Portuguese students, they found students most valued relationships with their peers and teachers and additionally stated that excessive workload and bullying were antithetical to their vision of happy schools (Gramaxo et al., 2023). Across the two studies, the authors reiterate the

importance of localization for implementation but found that the tenets of the Happy Schools Framework were largely valued in the Portuguese context.

Another study found that in cross-cultural comparisons there may be considerable overlap, but variation in what is most valued, attainable, or prioritized per culture should still be investigated. One study focused on such differences across two countries (Spain and England), but also in different populations within each (Lopez-Perez et al., 2022). At the national comparison level, primary school students in Spain (N=233) identified leisure time and communal harmony as qualities of happy schools. Results from the English students (N=421) found that positive relationships with peers and teachers and a sense of safety were most mentioned. The study also investigated the difference in what boys and girls in these samples identified as most important for a happy school, finding that girls mentioned relationships and emotional support more often than boys. These studies focus on what students have valued *most overall* in the sample. In other words, they provide support for the applicability of the framework through their findings that children in different countries also value what UNESCO has defined via Happy Schools. Their finding that students do not identify some of the characteristics in the framework in a qualitative survey does not imply those criteria should be removed. Instead, the study highlights that different cultures put different weight on how much they value a specific criterion.

In launching this new international version, UNESCO calls for multi-variable applications embracing the multidimensional definition presented by the Happy Schools project. This study combines numerous perspectives and several relevant constructs to empirically operationalize the Happy Schools framework. The present research is derived from the

conceptualization of the initial Happy Schools framework – the definitions are more extensive, and the criterion names are more accessible at face value.

International Frameworks and Assessments Address SDGs

The Happy Schools Framework operates within the field of international education accountability system that is interested in sustaining quality in schools beyond the straightforward preparedness assessment related to the SDGs. ILSA programs such as the Program for International Student Assessment (PISA) from The Organization for Economic Cooperation and Development (OECD) have developed and administered instruments to position their programs as responsive to complex international initiatives. What unifies these varied frameworks, projects, and goals is the demand that students be trained to navigate emotional and psychological states of being in themselves and others. A selection of these numerous endeavors is discussed here: PISA 2018 Global Competence Assessment, Assessment and Teaching of 21st Century Skills, PISA 2015, and PISA 2018 Wellbeing.

PISA 2018 Global Competence

As part of the 2018 administration of The Program for International Student Assessment (PISA), the Organization for Economic Co-operation and Development (OECD) developed a framework for Global Competence to measure students' ability to interact with the world outside what is familiar to them (OECD, 2019). The final framework consisted of four components across both cognitive and affective domains: knowledge of global/cultural issues or differences; skills (reasoning, adaptability, open communication, adaptability) to digest that knowledge into understanding; and attitudes (openness) and values (valuing cultural diversity) of respect towards other cultures to keep an open mind to multiple perspectives. Gathering data on the latter was determined to be beyond the scope of the PISA assessment. Being concerned about schools

sustaining peace-seeking, tolerant students is essentially insisting that students learn to have healthy relationships that do not negatively impact another's wellbeing or their own. Meant to respond directly to SDG 4.7, the assessment comprised a cognitive portion in which students engaged with and responded to various scenarios and a background questionnaire targeting students' attitudes relevant to living in a multicultural world.

The Global Competence framework had significant challenges and faced numerous changes on the road to being implemented. With each iteration, the assumptions and coverage changed from being separate from the SDGs to eventually being directly responsive to and supportive of "collective wellbeing and sustainable development' (OECD 2018). The first version was rejected by the PISA governing board, and the second was drafted by a group of representatives from only four countries. The cognitive portion was not field tested, and subsequently about a quarter of participating countries declined to administer at least one of the two portions of the assessment in 2018. Substantial criticism followed the Global Competence framework and assessment through these different phases, namely the international comparability of the conceptualization of Global Competence. Some claimed that OECD promotional documents did not align with the narrow definition and small, Western pool of experts responsible for the Global Competence Framework accusing the OECD of having "entertained a somewhat limited conversation in developing its definitional and measurement frameworks" (Ledger et al., 2019).

One study found that the scenarios in the cognitive portion were not authentic to possible experiences of the target age group (age 15), and supported the findings of a previous study that the instruments presented stereotypes in the stimulus for items targeting multicultural constructs (Chandir, 2022, Chandir & Gorur, 2021; Salzer & Roczen, 2018). For instance, many questions

ask about actions students take, presented in yes/no response formats, to support global wellbeing such as reducing energy consumption at home or boycotting products. Several students point to a highly relatable phenomenon for that life stage that their parents make those decisions for them, often for pecuniary reasons rather than environmental consciousness. Imagining students who would be able to make these decisions for themselves turns to another criticism of the measure: that any variation is measuring wealth and cultural priorities rather than individualized beliefs.

21st Century Skills

Similar to Global Competence, initiatives based on theories of 21st Century Skills (also called "skills for the 21st century" but henceforth abbreviated 21CS) are somewhat more concerned with specific skills necessary in a modern workplace, including critical thinking, collaboration, communication, problem-solving, and digital literacy (Care et al., 2018). A key undercurrent in all these skills is the idea that they are transferable in that they could be useful in many situations and learned in different contexts. Towards supporting the international comparability of 21CS frameworks, one study questioned the universality of developing "global citizens" to understand the proliferation of "transferable" skills identified in documents on national education websites (Care, Anderson, & Kim, 2016). Of the 102 countries surveyed, 76 national education websites identified non-academic skills to "equip students with enabling skills beyond literacy and numeracy" such as communication, creativity, critical thinking, and problem-solving (Care, Anderson, & Kim, 2016). There was variation in the integration of these skills into curriculum documents, and in how practically advanced the progression of skill development with only six countries identifying skills, integrating them into their curriculum,

and providing a developed skill progression. Nonetheless, 21CS are embedded in the education landscape worldwide.

21CS are also designed to be taught in schools, understood to be the best opportunity for students to learn the competencies that are most valuable in the society they will interact in as adults. One assessment project on 21CS defined its framework using four constructs: ways of thinking; ways of working; tools for working; and living in the world (Griffin and Care 2014). The development group also devised a progression of skills to inform item development and to support meaningful interpretation of the results. The drafted assessment tasks were piloted, administered in cognitive laboratories, and a larger field test was administered to about 1,000 students in six countries. Final items were selected based on item performance when calibrated with the Rasch model to estimate student ability parameters. Cutscores of student ability determined the boundaries between six ordinal levels of "knowledge building" in the progression of possessing or acting on 21CS aptitudes. Student-level responses could be reported to teachers using these cutscore benchmark definitions such that teachers could understand what skills students were ready to learn.

PISA Measurement of Wellbeing

PISA developed a multidimensional wellbeing framework first for the 2015 administration that combined objective (i.e. material goods) and subjective theories (i.e. self-evaluations) (Borgonovi and Pal, 2016). Five dimensions were measured in the student questionnaires: cognitive wellbeing, psychological wellbeing, physical wellbeing, social wellbeing, and material wellbeing. The cognitive domain encompassed content knowledge and "self-beliefs" such as enjoyment, self-efficacy, and motivation. Physical wellbeing was measured by five items regarding physical activity, exercise, and eating habits. Psychological wellbeing

was measured by five items covering overall life satisfaction, career expectations, academic motivation, and test anxiety. The latter three components of this portion are not well-aligned with the main theorizations of psychological wellbeing described later in this chapter and also have questionable face validity for international comparison. The social domain relied on scales of multiple items each and comprised school belonging, social learning experience, relationship with teachers, relationships with peers, and relationship with parents. However, the integrative framework does not lead to holistic reporting – results are given in alphabetical league tables by scale or item (OECD, 2017). There is no overall evaluation of wellbeing or the five dimensions.

PISA 2018 used the same five-domain definition of wellbeing and streamlined the reporting of relevant items and scales in four areas (OECD, 2019). For each of the four reported indicators (life satisfaction, feelings, self-efficacy and fear of failure, and growth mindset), league tables reported response percentages by country. The wellbeing framework for PISA 2018 also aligns itself more closely with the wellbeing psychology literature by grounding the study at the intersection of eudemonic and hedonic wellbeing theories. Once thought to be mutually exclusive, many current theorists understand wellbeing to be a combination of self-evaluation of the pursuit of pleasure (hedonic) and purpose (eudaemonic) (Diener et al., 2018). Expanding from the PISA 2015 reports, more context was given about the characteristics of students with different response patterns. For example, the percentage of students reporting positive feelings was reported by country along with group percent comparisons between gender, socio-economic status, and immigrant status. However, some claim that traditional descriptive reporting practices betray the multidimensional intent of the framework (Rappelye et al., 2020). The same study argues there is unilluminated bias in the PISA 2018 methods and reporting since

the definition of wellbeing relies entirely on student perceptions of individual wellbeing that may be deflated in collectivist societies.

Review of Wellbeing Literature

There are urgent calls for children to develop non-academic qualities and skills for operating in a social world. Wellbeing is not only well-represented in the academic literature but has an established presence colloquially. Thus, wellbeing could easily be conflated with wellness, happiness, thriving, or "living your best life" without the precise refinement of the construct. Concern for wellbeing was historically linked to physical health and safety (i.e. objective wellbeing) but has developed as a concept such that a multitude of factors are understood to relate to it. Among these conceptualizations are characteristics that benefit the student inherently to develop a sense of self and promote lifelong contentment. It is necessary to establish what should be sought after in the wellbeing of children, particularly in schools. The progress tracking guidance given by the UN, UNESCO, and the SDGs does not include a definition of a satisfactory school social environment or personal wellbeing to pursue.

The scope of this section is to summarize the main theories that support empirical measurement of wellbeing and how that is adapted to measure the wellbeing of students and school systems. First, subjective wellbeing, often aligned with theories of hedonic wellbeing, is introduced along with similar theories such as flourishing (Diener et al., 2010). Secondly, an introduction to psychological and eudemonic wellbeing and its measurement is provided (Ryff, 1989). Finally, positive psychology literature is summarized, serving both as an umbrella for these numerous theories of wellbeing and also for the influence on schools through positive education (Seligman, 2002). For each, empirically validated instruments are identified and examples of adaptations across diverse contexts are given.

Subjective Wellbeing

Subjective wellbeing (SWB) is understood to be aligned with hedonic wellbeing, one of two conceptualizations of wellbeing dating to Ancient Greek philosophy. Hedonic wellbeing is the state of satisfaction with life derived from pursuing pleasure (physical, emotional, mental) and avoiding pain (Diener et al., 2018). This school of theory is defined by the extent that happiness is present in an individual's life as captured by "people's cognitive and affective evaluations of their lives" (Diener 1984; Diener 2000). Understanding the extent of SWB is influential at the societal level in that happy individuals are likely to be more sociable and more productive (Diener, 2010). SWB is a self-defined state, based on the values one holds about how their affect should be rather than based on an established external evaluative framework (Diener, 1984). Its measurement is additive; if you have more positive affect than negative affect, then you are overall experiencing positive SWB. Aside from positive and negative affect, SWB is understood to also comprise a personal evaluation of either general life satisfaction or contextual satisfaction. All are understood to be instantaneously fluctuating experiences and the evaluation of feelings at an instant of time subject to the beliefs of the individual (Diener, 2000).

Measurement

Among the first instruments developed to measure SWB were single-item measures asking respondents to assign an overall score to their feelings of wellbeing (Cantril, 1965; Campbell et al., 1976). Measures of singular facets of wellbeing have been more extensively adapted to international student populations, including as part of the PISA 2018 Wellbeing Questionnaire (OECD, 2020). This style of measure has a risk of being contaminated by person variation not explained by variation in levels of SWB, such as mood and personality, test-form-based issues like item order, systematic cultural differences, and social desirability (Diener,

1984). Diener (2010) advises taking numerous measurements of individuals to get reliable readings of SWB rather than non-representative, moment-to-moment reactions to the person's experience. The principle of the "hedonic treadmill" can make it difficult to judge if instantaneous wellbeing is reflective of overall wellbeing; humans tend to adapt overall to changes that might make things immediately bad but fine after time (Brickman and Campbell, 1971).

Thinking to policy implementation, if SWB is most related to temperament it might be difficult to theorize interventions to directly address societal SWB, or to understand the state of SWB in a population sample. Single-item measures such as the Gallup World Poll (Crabtree, 2010) are still in circulation, even included as part of the PISA Wellbeing measures. But, while they have been adapted for adolescent populations and in international settings, some argue that cultural transferability is unproven against widely variant definitions of a good life lived (Holte et al., 2014).

Current measurement theory tends to favor multidimensional instruments (Marsh et al., 2019). Flourishing is one such theorization that promotes the pursuit of positive human experience across various aspects of life. The 8-item Flourishing Scale was developed to measure key components of a good life such as relationships, meaning, and positive outlook in numerous administrations to US college students (Diener et al., 2009). Generally, flourishing is more aligned with eudaemonic pursuits rather than the level of positive or negative feelings. This might suggest that flourishing is suitable for cross-cultural comparison. The Flourishing scale has also been adapted in some international contexts such as in adults in Iran (Hojabrian et al., 2018) and adolescents (11-16 years old) in Spain (Crous, Casas, Gonzales-Carrasco, 2018).

One study surveyed European adults on the extent of their "flourishing" as a multidimensional construct, particularly to see if it better captured international wellbeing than single-item measures (Huppert & So, 2013). Their approach to international comparability was to establish an internationally acceptable definition of depression and anxiety, then develop a definition for "flourishing" based on the opposite. Items were selected in the European Social Survey that corresponded to each of the ten features identified in the definition process. Ultimately, one item was selected to represent each of the features (shown in Figure 2.1) – the authors advise that having more items available per feature would have been more psychometrically sound.

Figure 2.2

Flourishing Definitions and Chosen Items From Huppert & So (2013)

Table 1 Features of flourishing and indicator items from the European Social Survey

Positive feature	ESS item used as indicator
Competence	Most days I feel a sense of accomplishment from what I do
Emotional stability	(In the past week) I felt calm and peaceful
Engagement	I love learning new things
Meaning	I generally feel that what I do in my life is valuable and worthwhile
Optimism	I am always optimistic about my future
Positive emotion	Taking all things together, how happy would you say you are?
Positive relationships	There are people in my life who really care about me
Resilience	When things go wrong in my life it generally takes me a long time to get back to normal. (reverse score)
Self-esteem	In general, I feel very positive about myself
Vitality	(In the past week) I had a lot of energy

The survey was administered to a representative sample of 43,000 adults in participating European nations. Exploratory factor analysis was conducted with and without a single-item measure for life satisfaction to investigate what features shared the most variance i.e. loaded on the same extracted factors. While two factors were extracted in the model without the life satisfaction item, the model including it extracted a third factor where only life satisfaction and

'positive emotion' were significant loadings. The authors concluded that their analysis indicates that flourishing is thus capturing more than life satisfaction alone. Further country-level analysis was also conducted to demonstrate relative strengths were not only concentrated at the top of the flourishing ranking. For instance, they highlight that France is in the middle of the rankings overall, but highest in 'engagement' and lowest in 'self-esteem'. In other words, their analyses suggest variation in the definitions of wellbeing, in that the highest overall performing nations were not the highest in all 10 of the proposed constructs. While they also rank countries based on an overall evaluation of flourishing, they advocate for more nuance in how wellbeing studies are reported. Reporting does not have to be limited by ranking, but can also expand on the unique variation of the participants. From this, they advocate for measures of wellbeing that combine the main theoretical schools – eudaemonic/hedonic or subjective/psychological. They make no claims as to what profile of wellbeing factors is satisfactory.

Psychological wellbeing

Psychological wellbeing (PWB) contends that the definition of happiness found in SWB is not based on theory, but instead on measures of convenience from studies not endeavoring to define the psychological structures of wellbeing (Ryff, 1989). This conceptualization of wellbeing is understood to be responsive to the eudaemonic definition of wellbeing, which emphasizes that positive human functioning is attained through the Aristotelian pursuit of meaning and purpose (Diener et al., 2018). Instead of evaluating the affective state, PWB is defined by desirable psychological characteristics that support a good life. Ryff (1989) is foundational in establishing the structural definition of "positive human functioning" to support the development of survey instrumentation to validate the structures.

Measurement

In the initial formulation of PWB, "points of convergence" from previous theories on wellbeing form the basis of six core dimensions of PWB (Ryff, 1989), defined in Table 2.2. The six-part framework led directly to instrument development. The first task was to write items for the six components based on agreed-upon descriptions of the maximum attainable positive and negative presence of the components. For example, a high scorer on the *positive relations with others domain* would be someone who "Has warm, satisfying, trusting relationships with others; is concerned about the welfare of others; capable of strong empathy, affection, and intimacy; understands give and take of human relationships" whereas their low scoring counterpart is described as "Has few close, trusting relationships with others; finds it difficult to be warm, open, and concerned about others; is isolated and frustrated in interpersonal relationships; not willing to make compromises to sustain important ties with others" (Ryff, 1989, p.1072).

Table 2.2

Foundational Definitions of Psychological Wellbeing

Core Dimension	Definition from Ryff (1989)
self-acceptance	holding positive attitudes toward oneself
positive relations with others	warm relating to others
autonomy	resistance to enculturation
environmental mastery	choose or create environments suitable to his or her psychic conditions
purpose in life	a sense of directedness, and intentionality
personal growth	that one continues to develop one's potential, to grow and expand as a person.

Ryff (1989) theorized this six-part framework was at least as good as describing wellbeing via existing alternative measures. This initial instrument with 32 items per scale was administered to a small sample of 321 adults alongside previously developed wellbeing

instruments, although a subsequent study had a larger sample of 1022 adults (Ryff & Keyes, 1995). All the new factors were significantly correlated, and all were positively correlated with measures of positive psychological functioning and negatively correlated with negative functioning. Principal components analysis extracted three factors, with most of the existing measures loading onto the first factor. Two additional factors accounted for 5% of the variance each and included the new indicators. Ryff claimed that the indicator-based model was more responsive to "the difference between feeling good at the moment and the more demanding task of realizing one's true potential" (Ryff 1989, p 1077). The later iteration conducted a confirmatory factor analysis on an 18-item version (Ryff and Keyes, 1995). This version has been successfully adapted to international samples of adults (e.g. Calderon et al., 2020; Kallay Rus, 2014; & Karas Ciecuch, 2017) and students between 10 and 18 years old (e.g. Viejo, Gómez-López, & Ortega-Ruiz, 2018; Stavraki et al., 2022; Strelhow, Sarriera, & Casas 2020).

Positive psychology

The field of positive psychology advocates for the development of positive health in the field of psychology - not just the correction of abnormalities - and prioritizes wellbeing as a desired outcome over emotional neutrality (Seligman, 2002). Focusing on what defines a positive condition rather than just the absence of a perceivable negative is influential in schooling too. Positive psychology has influenced the role of wellbeing in school through the field of Positive Education (Clarke 2020), promoting learning for global wellbeing and beneficial conditions "rather than only an absence of individual pain" (Hargreaves & Shirley 2021, p. 15). The five major components of positive psychology (positive emotion, engagement, relationships, meaning, and accomplishment, or PERMA) (Seligman, 2012) were developed into a psychometrically validated measure, the PERMA-Profiler (Butler and Kern, 2016) that has been

successfully adapted into numerous languages and contexts (e.g. Alves, Palemer, Gouveia, 2023; Marsh et al., 2019), including among school staff (Kern, Adler, Waters & White, 2015), undergraduate students (e.g. Giangrosso, 2021), and with secondary school students (Chue, Yeo, Nie, & Chew, 2023).

Special Considerations for International Student Populations

Much of the initial work towards these main theories for positive human functioning was conducted on adults and in single cultural settings. There have been validity studies to adapt versions of the instruments that have come from these foundational pieces that show promise for transferability across cultures and age groups. However, there are contentions in the literature that wellbeing in children requires additional theoretical care. There is no question that children can distinguish between good and bad feelings (Schonert-Reichl et al., 2013). However, there is risk in misrepresenting student wellbeing as a point of validity but also as a point of real-world consequence. Population-level studies can have significant influence on policies that impact students in situations much different from the summary value of that aggregated measure (Kelly et al., 2020). Students at the fringes of societal distributions are disproportionately impacted by both of these concerns.

In response to these considerations, it becomes evident that sustaining wellbeing in students is highly sensitive to environmental stimuli. Finally, definitions of wellbeing must both be grounded in theoretical legitimacy and valued in target cultures to be meaningful across populations.

Cross-cultural Comparisons

This research aims to support complex evaluations via the multidimensional definition of school climate health; it is necessary to directly face the challenges of conducting this research

for an international dataset. Evaluations rely on valid arguments grounded in theory specific to a particular context and responsive to evidence gathered. These components operate together only in a certain context – thus a multicultural evaluation can only be valid if the argument is valid across contexts. While transferability in the context of the present research refers largely to cultural differences across entities, culture is critical to context at any scale. In any group, "culture refers to a set of beliefs, attributes, values, knowledge, and skills that collectively creates identity." (Kirkhart, 2010). Without proper consideration for context at all stages, the results of any evaluation are of limited use at best or possibly to the detriment of stakeholders. Furthermore, the final results of the evaluation must acknowledge the risk to the validity of extracting conclusions from the appropriate context. (Fitzpatrick, 2012). Assuming an evaluation stance in this research is appropriate to put consideration of culture and value at the forefront of the analytic framework and rubric to be created.

As mentioned in the discussion around SWB, wellbeing is "socioculturally formed," (Tov and Diener, 2009), which complicates the evaluative task of establishing agreed-upon benchmarks. While some studies argue that a basis for similarity exists, (eg. Krys et al., 2023) one study points to four problematic assumptions often overlooked in studies comparing subjective wellbeing across cultures: wellbeing is not equally valued in all cultures; prioritizing individual wellbeing may not be universal; that a wellbeing measure works the same across cultures; and wellbeing might be inseparable from academic achievements in some cultures (Clarke, 2020). Some existing large-scale measures of wellbeing do not adequately integrate different components of wellbeing towards holistic measurement, even when based on a holistic framework. The PISA definition of wellbeing was multidimensional but the instrument and reporting were individualized, and key dimensions were measured by a single item (OECD,

2020). For some critics, this implies a single unified view of life satisfaction and shows a lack of attention to the literature and cultural diversity (Rappeleye et al., 2020). Allowing for both a priori and emergent definitions could support meaningful reporting of divergent presentations of wellbeing.

Many studies pointing to issues of cross-cultural comparison interpret differences that exist as the result of potentially biased instruments rather than interrogate different qualifications on how wellbeing, happiness, or satisfaction are defined across different cultures. When measuring wellbeing, the forces of cultural practice and norms could account for the majority of variation in wellbeing levels. Numerous distinctions can be made when comparing national cultures, but among the most important separating paradigms of international measurement of wellbeing is between individualist and collectivist cultures (Joshanloo et al., 2021). These demarcations also capture common geographic or geopolitical groupings; Western or European cultures are more individualistic than Eastern or Asian cultures, although it is important to keep in mind that generalization is less interesting than the diversity within these groupings. One study comparing Russians and Americans found more reticence to express happiness among the Russians but found that reluctance was unrelated to their wellbeing (Sheldon et al., 2017). Cultures that strive for hedonic wellbeing prioritize choosing positive feelings whenever possible to improve one's condition whereas in others the pursuit of self-enhancement is not so strongly endorsed (Rosenman & Kurman, 2019). Joshanloo and colleagues (2021) argue that the cultural influence in defining what is valuable in all areas of life explains the variance seen in wellbeing measures across cultures.

Two defining cultural spheres can be further understood as cultures where individuals "see themselves as unique, promote their own goals, and seek self-expression" and those where

individuals "seek to belong and fit in, to promote others' goals, and to occupy their proper place" (Berry, Poortinga, Segall, & Dasen, 2002, pp. 101–102). For cultures of the latter type, an eudaemonic approach is more appropriate, so attempting to balance wellbeing theories is not only reflective of the literature but also suitable for international comparisons. SWB has been found to be strongly correlated with GDP, with some authors concluding such life satisfaction measures are indirect measures of economic development (Diener & Tay, 2015). Studies have also found that measures of eudaemonic wellbeing correlate less strongly with GDP (Joshanloo, 2018) in addition to individualistic values of self-promotion (Krys et al., 2019). While so-called developed countries may score highly on SWB measures, other countries may outperform on eudaemonic measures that better match how those countries value collective wellbeing. Oversimplifying either construct to represent a unified definition of life well lived is theoretically and empirically problematic. Further, being able to distinguish "self" is also culturally dependent:

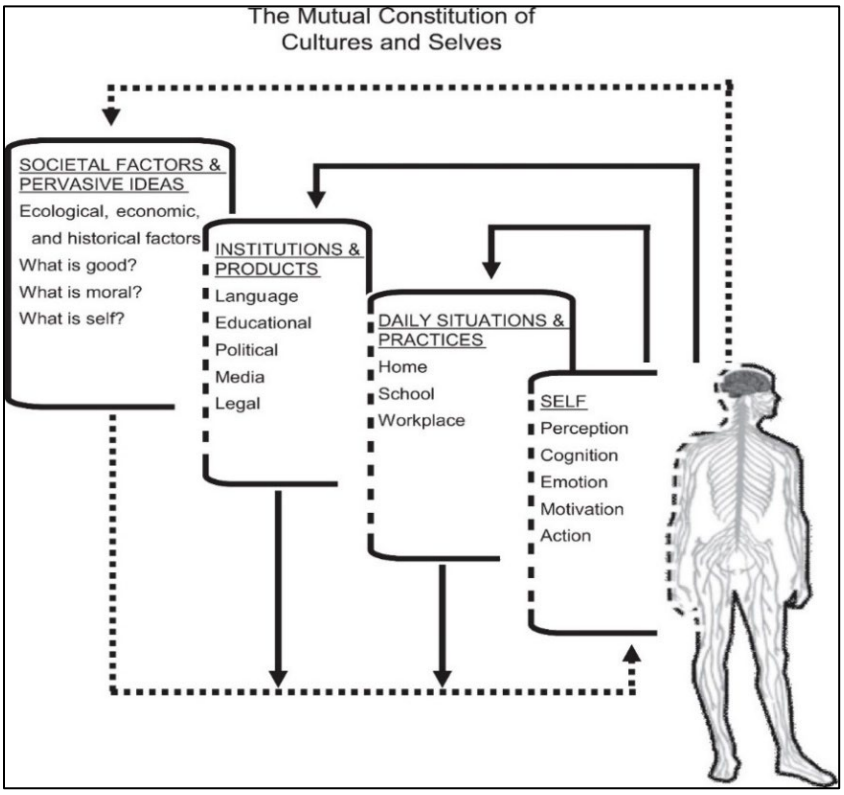
> When deciding how satisfied they are, people in individualistic nations find it natural to consult their affect, and feeling pleasant emotions frequently is a reasonable predictor of life satisfaction in these societies. In contrast, people in collectivist cultures tend to more often consult norms for whether they should be satisfied and to consider the social appraisals of family and friends in evaluating their lives. (Diener, 2000, p39).

The theory of mutual constitution of self and culture is relevant to introduce to understand this inherent methodological concern. As shown in Figure 2.3, the different contexts that an individual exists in define and support each other (Uchida and Rappeleye, 2024; Markus and Kitayama, 2010). Thus, an individual responding to a self-report survey asking questions

related to their satisfaction could target completely different perceptions based on their cultural answer to 'What is good/moral'.

Figure 2.3

Visual Representation of Mutual Constitution from Markus and Kitayama (2010)



Inclusive definitions and appropriate specificity are essential to targeting a conceptualization of wellbeing that can be compared over many cultures – and empirical studies have investigated the impact of these differences on evaluating wellbeing. Suh, Diener, and Updegraff (2008) hypothesized, that members of highly individualistic cultures (e.g. United States) would access emotional perceptions to evaluate their life satisfaction (i.e. their positive or negative feelings) whereas those in communalistic societies (e.g. East Asia) would be more sensitive to the *cultural appropriateness* of being satisfied. The distinction of the Suh, Diener, and Updegraff (2008) study was the theorization and testing around the mediator that would

explain this variation. Building on theories of interdependence (Markus and Kitayama, 1991), the study aimed to test the variations in how self-construal, the extent to which and how individuals self-define uniquely or to groups and norms, correlated with a component of SWB – overall life satisfaction.

Regression analysis was conducted to examine the relationship between social approval and positive affect on life satisfaction self-report in a small sample (N=102) of American college students (Satisfaction with Life Scale (Diener, Emmons, Larsen, & Griffin, 1985); international adaptations include Gongora 2014; de Carvalho, Aquino, Natividade 2023; Gander, 2020; Dirytze, Perminas, Biliuniene 2021; Crous, Casas, Gonzalez-Carrasco 2018). The students also responded to a survey that would categorize them as individualist or collectivist (also referred to as "allocentric"). Results of the regression analysis found that life satisfaction scores were statistically significantly predicted by emotional affect, but that for allocentric individuals the extent that friends or families approved was also statistically significant. A second study was conducted with the same instruments but with the sample being randomly selected to be primed for individualist or communalist perspectives before responding to the surveys. The results were parallel to those in the first, although in a sample of Korean students that was part of the second study, *only* family or friend approval was predictive of life satisfaction in the group primed for collectivism. In all cases, the self-evaluation of collectivist individuals was reflective of a different cognitive process than more individualist respondents. Taken to cross-cultural comparisons, a narrowly defined life satisfaction instrument may not be unidimensional in capturing life satisfaction across contexts – some cultures may bring more to the question than personal emotion alone.

Suh et al. (1998) aimed to test this phenomenon at the national culture level in two studies with a total sample of 62,446 across 61 countries. Using the definition of subjective wellbeing as a combination of self-evaluations life satisfaction and balance of positive and negative affect, they hypothesized that the magnitude of the relationship between affect and life satisfaction would be higher in more individualistic countries than in collectivist countries. The first study was conducted with data from 55,666 participants (ages 16-99) in 41 countries that administered the World Values Survey II (WVSII, 1994) which included the Affect Balance Scale (Bradburn, 1969) containing five questions each on positive and negative affect and one life satisfaction question. Each country was given a rating from 1 (most collectivist) – 10 (most individualist) on an individualism-collectivism spectrum based on the average of scores given by two contemporary scholars on the subject (Hofstede, 1980; Triandis, 1996). Their findings were that all countries had a positive relationship between affect and life satisfaction (mean $r = 0.41$) but that comparing the correlation coefficients (via conversion to Fisher's z-scores) showed statistically significantly higher coefficients in individualist countries than in collectivist. While they were unable to test the influence of societal norms on conformity in the same study, the authors conclude their findings suggest "inner emotional feelings play a more significant role in one's judgment of overall life satisfaction in individualist nations than in collectivist nations" (Suh et al., 1998, p. 486).

Suh et al. (1998) conducted a parallel study among international college students (ages 18-25) from 40 nations. In addition to collecting data regarding affect and life satisfaction as in the first study, the study aimed to test the hypothesis that *societal norms* of life satisfaction are more highly correlated with life satisfaction in collectivist countries than in individualist countries. Suh and his colleagues included a section on the questionnaire that asked students to

complete the Satisfaction with Life Scale (Diener et al., 1985) once as a self-evaluation, then a second time they were instructed to complete the survey "how the ideal person should answer these items" (Suh et al., 1998, p. 487). The second study was designed to replicate the WVSII study, but also extend understanding of the varying definitions of subjective wellbeing in cross-cultural contexts. They found similar results as the WVSII study, although they add that their findings suggest more relevant differences in "how much [countries] regulate negative emotions than of the degree to which they treasure positive emotions" after comparing the correlation of negative affect and life satisfaction and that between positive affect and life satisfaction (p. 488). To address their hypothesis regarding social norms, Suh and colleagues conducted regression analysis, regressing affect and norms on life satisfaction for country group designations of "Individualist" "Collectivist" and "Neutral". Interpreting the standardized beta of each of the three models, the contribution of norms in predicting life satisfaction was much lower in individualist countries ($\beta = 0.161$) than in collectivist ($\beta = 0.345$), while emotions were more predictive of life satisfaction in individualist countries ($\beta = 0.556$) than collectivist ($\beta = 0.342$). Figure 2.4 is included from this study to visualize the differing proportion of variance explained by emotions and norms across the different country groups. Suh and colleagues conclude that the basis of judgments i.e. what is being stimulated by questionnaire items differs across countries based on these well-known philosophical differences.

Figure 2.4

Graphical results, Suh et al., (1998)

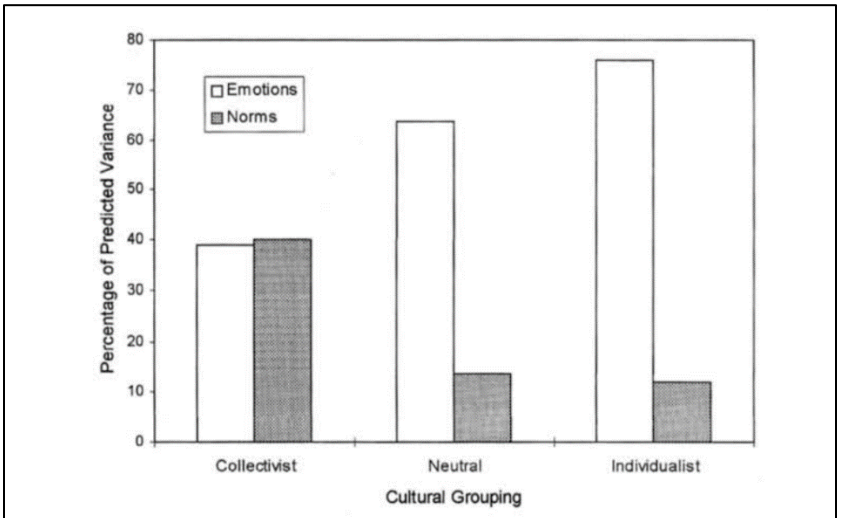


Figure 1. Relative contribution of emotions and norms in predicting life satisfaction across cultures. For the collectivist group, n = 2,210; for the neutral group, n = 2,097; and for the individualist group, n = 2,449.

Stepping back to take these studies as a whole picture, wellbeing measures are complicated to apply comparatively across multiple contexts. Not only are there national cultural concerns, but individual-level expressions of those concerns (or unrelated personal tendencies) are more complex than a simple alignment around a single self-concept evaluation. Considering the assessment of such a construct as wellbeing, the cognitive process being targeted by the questionnaire stimulus must be comparable for all participants, or the evaluation must embrace the complexity at the individual and group levels. However, these studies do not suggest that global wellbeing is impossible to study, but that the Western psychology perspective that emotions are the most important predictor of evaluations of life satisfaction needs to be questioned as an underlying, nontransferable assumption. Nonetheless, large-scale initiatives such as PISA still rely on comparing self-report information to single-item wellbeing measures. In their discussion, Suh and colleagues hoped instead that there would be advances in understanding the "unique determinants and correlates of SWB within individual cultures" (p.

491). This dissertation hopes to disrupt the tradition of generalizing a prescriptive definition of wellbeing onto participating countries, embrace emergence, and focus on meaningful interpretations relevant to student life in context.

Multidimensionality

Following directly from the literature on the complexity of wellbeing and the international variation in understanding a 'good life', reductive lenses to measuring the wellness of a community are inappropriate. Contemporary theory affirms that wellbeing is both hedonic and eudaemonic, and not fully measured by affect alone (Adler and Seligman 2016). Furthermore, measurement by subdomains in the PWB fashion can better inform what aspects are most important for wellbeing and achievement alike, lending itself also to studies of environmental indicators of wellbeing (Clarke, 2020; Jayawickreme et al., 2012). Measures that target multiple subdomains of wellbeing are both more comprehensive summaries of human experience and provide workable data in policy contexts (Forgeard, Jayawickreme, Kern, & Seligman, 2011; Kern & Wehmeyer, 2021). Wellbeing is contemporarily understood to comprise thriving across multiple dimensions in life and should be treated as a dynamic process (Adler, Seligman, 2016). As a dynamic process, it can be influenced by the circumstances of life in addition to mood or personality, requiring repeated measurement over time if a single dimension of wellbeing is used. While multidimensional measures also benefit methodologically from repeated measures, the crucial implication here is that single-dimensional measures of wellbeing are especially unreliable as reliable measures of such a complex construct. A well-defined multidimensional measure would at least give a more comprehensive picture of a complicated state of being.

An additional instrumentation issue in measuring the dynamism of wellbeing in children at school is the role school plays in the lives of children. Measuring SWB in a school setting may result in deflated measures – schools are "anti-hedonic" in that children must undertake tasks they would not choose for themselves but that could support student's lifelong learning (Clarke, 2020). To understand how well students feel in school it is key to understand wellbeing as a social issue of "inclusion, belonging, peacefulness, and human rights" (Hargreaves & Shirley, 2021, p.5). A socioecological perspective is encouraged, to test how the qualities of key components impact wellbeing and impact the relationship between wellbeing and academic achievements (Allen et al., 2022). To seek effective interventions, it is necessary to understand the coherence of different tiers of influences that impact wellbeing (Vella-Brodrick et al., 2022). Ultimately, students are well at school to the extent that their perceptions of that environment support beneficial development (Clarke, 2020). A problem with measuring school wellbeing in students based on just their self-reports is that school itself is a short-term nuisance for long-term benefit. While children are capable of self-interrogating their feelings and state of being to respond to such questions (Schonert-Reichl et al., 2013), unidimensional measures of wellbeing might measure how much fun the student feels they are having rather than how well they are. This orientation does not provide clear inroads for constructs malleable and impactful enough to make a change that would alter a student's state of wellbeing

Considering the school environments that support wellbeing is essential to promoting student wellbeing internationally. While theoretical understanding of wellbeing has unified around a multidimensional vision of living life, the complications of measuring wellbeing in schools and around the world suggest that another avenue for understanding the expansive construct may be necessary. Furthermore, the intervention potential to improve student

wellbeing directly is limited if one does not also have reliable data on the likely influencers of a student's wellbeing. If the ambition is to measure student wellbeing and then improve on it, it is not possible to directly inject children with wellbeing alone. It would be most efficient to be able to capture how well a school is functioning as a supportive social system at the same time as how well they are likely supporting the wellbeing of students and adults alike.

School Climate and a Systems Approach

Defining School Climate

The conditions of a school's environment represent a combination of interrelated factors at multiple levels that influence the actions of those involved, rather than physical or administrative school structures. An extensive review of school climate literature from Thapa et al., 2013 defined school climate to be composed of safety, relationships, teaching and learning, institutional environment, and school improvement processes. (See Table 2.3 for elaboration). The most recent framework from the National School Climate Center (NSCC) expands each of these to include several subdomains (also in Table 2.3). People at school act out varying roles that are either supported or corrected to define the overall school climate conditions. For example, school climate can foster exclusion by ignoring problematic behaviors, enabling bullying activities to be perpetuated (Riley, 2019). A productive school climate is characterized by members' caring relationships and meaningful and influential engagement, staff collaboration and continuing education, supportive administration, staff and student peer mentoring, active parental involvement and education, involvement of various stakeholders, and the participation of the local community (Cefai and Cavioni, 2015).

Table 2.3

Select Definitions of School Climate

Thapa et al., 2013	NSCC, 2023
Safety	Safety 1. Rules and Norms 2. Sense of Physical Safety 3. Sense of Social-Emotional Security 4. Online Safety
Relationships	Interpersonal Relationships 5. Respect for Diversity 6. Teacher-Student Relationships 7. Peer Relationships
Teaching and Learning	Teaching and Learning 8. Support for Academic Learning 9. Social and Emotional Learning
Institutional Environment	Institutional Environment 10. School Connectedness 11. Physical Surroundings 12. Social Inclusion
School Improvement Process	Leadership and Efficacy 13. Administration and Leadership 14. Collective Efficacy

Schools and schooling systems are complex networks of human actors with relationships that can support or diminish student academic success and emotional wellbeing, largely orchestrated by the adults involved (Ramelow et al., 2015; Lombardi et al., 2019). Studies have also attempted to capture what is most important to the school climate for students. Most recently, Konishi et al. (2022) conducted qualitative interviews with a small sample of secondary school students (N=22) in Canada to test a priori hypotheses of dimensions of school climate most important to students. In addition to these dimensions, the most commonly named were school safety, school belonging, and adult acceptance of diversity, derived from student responses. Although not anticipated by the authors, the most frequently identified dimensions of these were peer interaction, student-teacher relationships, school resources, and school

order/disorder. Intending to develop metrics of wellbeing in schooling systems, Lee et al. (2017) conducted a confirmatory factor analysis from data collected from Australian middle school students (N=7209). The measure consisted of 32 Likert items, conceptualizing school climate as four subconstructs: student-student relations, student-staff relations, academic emphasis, and shared values & approach.

Returning to a previously introduced example, the Middle Years Development Instrument (MDI), Australia administered an adapted version of the MDI to a substantial sample (N=51,574) of 10-15-year-old students over two years finding that peer and teacher interactions they experience as key to the conditions of school climate, autonomy, and fair discipline, along with freedom from bullying. The preparedness of students and teachers to engage in learning was also important to sustain academic and community support, and also school order/disorder (Gregory et al., 2019). Although this study found the instrument had satisfactory psychometric qualities, the instrument has not been tested outside of Commonwealth countries.

Relational Trust

The qualities of relationships have long been associated with effective school change and productivity. Bryk and Schneider (2002) found what they defined as relational trust in school communities was the most predictive of school success, including achievement, beyond the demographic factors of Chicago schools. School systems in their research functioned based on the discernment of people operating in the school. Based on the obligation of a certain role, a person operating in the system would interpret (i.e. discern) meaning in the other's action to inform their reaction. They argue that "these judgments of intentionality are grounded in each individual's historical perspective on the institution, personal and cultural beliefs rooted in his or her family and community of origin, and prior workplace socialization experiences" (Bryk &

Schneider, 2002, p. 21). Further, relational trust accounts for the influence of relationships among different stakeholders in the school; student-teacher, principal-parent, teacher-principal, and teacher-teacher.

The relationships that are modeled for students at school, and how school makes them feel about themselves, can impact students' lifelong strategies for engaging with others. While critical peer relationships are not included for students, the definition of relational trust is multidimensional and cognizant of how power influences interactions in context. These dynamics impact the discrete interactions of people in the system by interacting with the large-scale perceptions of the school system, which they define as the extent that there is trust among the various roles in school. The magnitude of relational trust in a school community influences the mechanisms that sustain the positive outcomes of highly functioning schools: "Networks with high levels of trustworthiness maintain socially desirable norms and sanction unacceptable action" (Bryk & Schneider, 2002, p. 14). Relational trust influences the discernment of others in the school system; actors rely on the respect, competence, personal regard for others, and integrity of the school when interpreting the actions of others. If relational trust is established in a school, then the social system could self-sustain positive environmental qualities to the benefit of all involved. This foundational work is included here not only as a reference point in school climate literature but also as evidence that theories of schools as dynamic systems are supported by longstanding empirical studies.

Ecological Systems Theory

With attention to interrelatedness, multidimensionality, and complexity, it is not surprising that many recent studies have taken systems thinking into their theorization, most commonly, the work of Bronfenbrenner (as in Bronfenbrenner, 1989) in establishing socio-

ecological theories for human social systems. Also referred to as the bioecological model, this framing of human development and personal potential stresses the importance of the outside influence of circumstance on children (Waughn & Guhn, 2014). With a bioecological model of a school system, school climate could be defined as "the affective and cognitive perceptions regarding social interactions, relationships, safety, values, and beliefs held by students, teachers, administrators, and staff within a school" (Rudasill et al., 2018, p. 46). Further, Allen et al. (2018) propose that school belonging be reconceptualized with a socioecological lens to explicitly define the mechanisms operating at different levels of a school social system. It is essential to understand systems as a combination of interrelated parts that operate holistically within the boundaries of their context (Bronfenbrenner, 1992). A system in a state of wellbeing supports both individuals and the collective towards efficiently pursuing the shared goals of the system. In theorizing the desired state of social systems Jenks (2004) proposed that the wellbeing of such systems is supported by three components: trust, agency, and opportunity. In the pursuit of sustainable, positive education systems, it is important to understand what types of educational environments are most supportive of that development.

Looking at the intersection of the multidimensional theories of wellbeing already discussed also points to factors that align with major components of school climate. There is substantial support for aligning the needs of international evaluations with the theories of wellbeing and school climate. Contemporary theory on wellbeing advocates for multidimensional measures of various qualities in life. From bioecological theory, the systems that impact human life are systems made of human interaction defined by roles and expectations (Bronfenbrenner, 1992). Some measurement research on school climate has explicitly combined aspects of wellbeing, school climate, and systems thinking to capture a holistic understanding of

school-level wellbeing. One study validated an observational instrument to measure "flourishing learning environments" in K-12 classrooms in Texas (Waxman et al., 2016). The Flourishing Classroom Observation Measure was designed to examine five components of what the authors defined as flourishing: a) positive emotion, (b) engagement, (c) meaning, (d) positive relationships, and (e) accomplishment (Waxman et al., 2016, p.3). These factors combine the positive affect component from SWB and hedonic theories of wellbeing with several factors from PWB in the context of describing school climate.

In a sample of 988 Canadian university students, Zandlivet et al., 2019 developed and administered The Healthy Environments and Learning Practices Survey (HELPS) to validate a multidimensional configuration of satisfaction with learning environments. The instrument was developed to understand "how psychosocial aspects of learning environments…can impact student wellbeing" as measured by flourishing (as in Diener et al., 2010 and Seligman, 2011) to define classroom wellbeing (Zandlivet et al., 2019, p.284). Among their findings was that classroom environments could be distinguished and that "as students' perceptions of the learning environment become more positive, so, too, do their self-reports of life satisfaction." (Zandlivet et al., 2019). Relationships and their qualities present themselves as a powerful unifying factor across these dimensions, acting as an important component of positive wellbeing individually and in the school climate. For reference, Table 2.4 is included to summarize all instruments mentioned in this chapter, along with a selection of relevant adaptations.

Table 2.4

Overview of Referenced Instruments

Name	Original Authors	Methodology	Sample
Flourishing in Europe	Huppert & So, 2013	Quantitative; secondary analysis	
HELPS	Zandlivet et al., 2019	Quantitative; validation study	988 university students; Canada
Flourish Classroom Observation Measure	Waxman et al., 2016	Quantitative; validation study	
Happy Schools Frameworks	UNESCO, 2020	Qualitative; open-ended survey	

				Adaptations		
Name	Original Authors	Methodology	Sample	Country	Sample	Reference
Satisfaction with Life Scale	Diener et al., 1985	Quantitative; instrument development		Argentina Brazil Germany Lithuania Spain	247 13-18 yos 1327 adults 600 university students 2003 adults 763 11-16 yos	Gongora 2014 Carvalho, Aquino, Natividade 2023 Gander, 2020 Dirytze, Perminas, Biliuniene 2021 Crous, Casas, Gonzalez-Carrasco 2018
MDI	Shonert-Reichel et al., 2013		51,574 10-15 yos. Canada	Australia	51,574 10-15 yos	Gregory et al., 2019
Psychological Wellbeing	Ryff and Keyes, 1989; 1995	Quantitative; instrument development	321 adults; 1022 adults. Both US	Thailand Spain Romania Poland Brazil Spain	1122 university students 330 elderly 664 adults 2035 13-78 yos 1248 12-18 yos 1590 13 to 19 yo	Calderon et al 2020 Granados et al 2023 Kallay & Rus 2014 Karas & Cieciuch 2017 Stelhow Sarriera Casas 2019 Viejo, C., Gómez-López, M., & Ortega-Ruiz, R. (2018)
Flourishing Scale	Diener et al., 2010	Quantitative; instrument development	689 university students. US	Iran US Spain	412 adults 1,035 adults 763 11-16 yos	Hojabrian et al 2018 Marsh et al., 2019 Crous, Casas, Gonzalez-Carrasco 2018
PERMA-Profiler	Seligman, 2002; Butler and Kern, 2016	Quantitative; instrument development	31,966 adults; international	Brazil Italy Greece Portugal Australia Singapore	1327 Adults 2533 university students 331 adults 17-67 1258 adults 515 high school students; 143 school staff 3788 11-14 yos	de Carvalho, Aquino, Natividade 2023 Giangrasso 2021 Miskidou et al., 2021 Alves, Palemer, Gouveia, 2023 Kern, Adler, Waters & White, 2015 Chue, Yeo, Nie, & Chew, 2023

Considering Achievement and External Factors

Achievement, Wellbeing, and School Climate

UNESCO posits that Happy Schools improve the welfare of everyone involved, which will lead to more productive learning environments (UNESCO, 2024). While the merits of a healthier school climate (safe, productive, supportive environments for students and adults alike) are self-evident at the surface level, understanding any impact on academic achievement is relevant to policymakers. However, the body of literature studying the relationship between achievement and wellbeing or school climate is not clear-cut. Complications exist even in the fundamentals of what measures are compared - across studies and contexts, the exact measure representing "achievement" varies from GPA to teacher evaluation, to large-scale measures (Kaya & Erdem, 2021).

There is support in the literature that schools can promote improved academic outcomes and prepare children to lead satisfying lives at the same time. Wellbeing is assumed to be a skill that can be taught, and a possible mechanism for social progress initiatives at large scale (Adler and Seligman, 2016). It was concluded that adequate social-emotional skills like wellbeing can help students with academic resilience and improve their academic experience (Cefai and Cavioni, 2015). There is room for schools to both sustain good academic environments and good wellbeing, and student wellbeing is not isolatable from the impacts of the child's situation (Clarke, 2020). Social relationships are key to how people evaluate their overall life satisfaction over the course of their lives (Diener 2018). Even before the recent effects of the COVID-19 pandemic, studies found that life satisfaction in students was declining in most countries internationally (Marquez and Long, 2021). Thus, it is dangerous to make claims on the relationship between wellbeing and achievement "without broader awareness of the respective

school, home and societal ecosystems in which they are nested." (Clarke, 2020, p. 284). Children are part of a social ecosystem when they are in school and are also influenced by their society and culture, which also impacts their academic outcomes.

Further to the theoretical framing of this research, three areas that complicate the relationship are the complexity of wellbeing as a construct, the social system that defines the state of school climate, and the multitude of external factors that impact the attainment of both wellbeing and school climate.

The nature of the first issue has been explored already in the literature review of this dissertation – different schools of thought and approaches to the construct, as well as cross-cultural variation, contribute to the complexity of extracting a unified definition of wellbeing. Thus, researchers interested in the relationship between wellbeing and achievement often focus on a subset of the overall construct, such as focusing on psychological wellbeing alone for their study (Amholt et al., 2020). Seemingly contradictory results cloud a clear understanding; careful attention must be paid to the definition of wellbeing and the treatment of common confounding factors.

A recent meta-analysis study compiled and synthesized 77 studies concerning wellbeing and achievement to attempt a holistic understanding of the relationship (Kaya & Erdem, 2021). Among their findings, they found that the magnitude of the positive relationship between wellbeing and achievement was greatest in elementary schools than in other academic settings. However, there was significant variation in the magnitude of the relationships depending on publication year, studies in peer-reviewed journals, and comparing 'levels of human development' across countries and wellbeing domains. Overall, the authors conclude that the

literature reviewed supports “the theoretical basis” for a positive relationship between academic achievement and wellbeing.

Addressing the second complication, school climate in this paper has already been introduced as a complex social network comprising several levels of students and adults that have interconnected roles to sustain school functioning. Several in-depth qualitative studies have found that for school-age children, the most important factor in their estimation of wellbeing is their relationships with peers and adults (Vjucic et al., 2019; Konishi et al., 2022). Echoing Bryk & Schneider 2002, the qualities of relational trust at work among the students, teachers, parents, and school leaders are most indicative of the state of the overall school atmosphere. The extent that the adults in schools can sustain their wellbeing is also indicative of their job satisfaction and decision to stay in their jobs (Dreer, 2024). These different levels influence and play off of each other such that focusing on a subset alone is not comprehensive.

Many studies investigating the relationship between achievement and school climate employ only student-level measures of the construct (Berkowitz et al., 2017). Considering the established theorization of schools as social systems, these methods neglect how adults evaluate the environment. While student perceptions of school climate have been shown to relate to academic achievement, studies have found that teacher evaluations were more attuned to shifts in classroom-level climate (Mitchell et al., 2010; Wang and Eccles, 2013). Previous studies synthesizing numerous perspectives on school climate through multilevel modeling have found that the perceptions of staff and student perceptions of school climate are both predictive of numeracy achievement (Maxwell et al., 2017). Relying on student perceptions alone may dampen the positive qualities that the overall community is experiencing. Given the importance

of teacher-student relationships in how students and teachers perform their roles in school, only considering student perspectives leaves part of the system underexplored.

Finally, any research into the relationship between achievement and wellbeing or school climate risks oversimplifying the role played by numerous contextual factors that act on children both inside and outside of school. It is crucial to investigate this relationship with a broad, integrative framework of the social systems of student and school life (Clarke, 2020). For instance, cultural differences within the same student body could confound such conclusions. A study investigating achievement and wellbeing among multilingual students found that the composition of the student body had an impact on the magnitude of the relationship. Namely, multilingual students in more linguistically diverse schools compared to those in less diverse schools felt more supported and teachers were more active in adapting lessons to their needs (van der Wildt, 2017). Creating a classroom or school environment that is conducive to diverse students is an intentional practice that can be trained. Thus, in the absence of such intention, evaluations among students exposed to the same environment could be vastly different via the extent of acceptance being fostered. Positive classroom and school environments are the result of "the intentional design of positive social interactions between and among groups" to sustain a productive and safe learning environment (Hymel and Katz, 2019, p. 337). Ultimately, numerous factors at home, at school, and among their peers impact whether the psychological state of wellbeing can be sustained for students, and instrumentation and methods that seek to narrow the lens risk assigning blame to the most in need of support.

Considerations for External Factors

The role of these external factors in this research is to explore the definition of the school climate health clusters based on variables important to schooling but not related to attitudes or perceptions. Even just within the availability of TIMSS the selected variables are not intended to be exhaustive of those which could interestingly define the schools more likely to be in each group. However, they are critical and commonly used in the literature such that their inclusion here contributes to the results of the study and may have implications for further research.

The variables chosen for secondary analysis are of interest in their potential impact on both school climate and wellbeing and on the relationship between these and academic achievement as measured by TIMSS. All variables are taken from the school questionnaire to reduce aggregating data further. The selected variables target either descriptive school characteristics or student demographics. There are numerous possibilities for characteristics of interest for these analyses, but there is the limitation of what is covered by the TIMSS context questionnaires and the feasibility of one study. However, there is support in the literature for the variables chosen: school urbanicity, school size, the proportion of students ready for instruction, the proportion of students from disadvantaged backgrounds, and the proportion of students speaking a different language at home than at school. It is also true that the variables and their importance to education research overlap. For instance, urban schools tend to be large and have high proportions of disadvantaged students so research that investigates one of these facets often also includes conclusions about another.

School surroundings. School urbanicity has been shown to impact achievement and comparing achievement across different types of school locations has been of interest. Observationally, being in a big city has allure if that city is well-regarded but less desirable if

there has been extensive suburban expansion that has drawn resources and population from the centralized location. TIMSS data has also been used directly to study the relationships between school surroundings and achievement (Tomul et al., 2021). However, the fact that the school is in a certain location is not the mechanism for variation. As with most demographic information, location implies a differential flow of resources or environmental features that influence school environments and outcomes. Further, school location alone does not prevent the influence of additional factors. For instance, one study of suburban schools found that it was still true that surrounding crime and student behavioral issues were as predictive of lower achievement as one might expect in an urban school (in the US context) (Sulak, 2016). Well-studied mediators of academic achievement also have been shown to be of different strengths depending on the urbanicity – higher SES was shown in a study of 9350 eighth grade students to be less strongly predictive of academic achievement in rural schools compared to urban (Miller and Votruba-Drzal, 2015). Context and knowledge of individual cases are crucial to any claims that make inferences about the quality of schools or students.

School size. Variations in school organization make the claims of school size hard to generalize as some studies interchangeably use school size and number of students per grade (Gershenson & Langbein, 2015) while others endeavor to understand the difference broadly between small, medium, and large schools (Lee & Loeb, 2000). Gershenson and Langbein used a longitudinal design and found on average there was no relationship between school size and performance. However, for schools with high proportions of disadvantaged students or students with learning disabilities, an increase in school size was found to decrease mathematics achievement. Lee and Loeb studied 264 schools of varying sizes in the US and found that small schools (less than 400 students) had higher mathematics scores than medium or large schools and

teachers had more positive attitudes. Some studies have found that the relationship between school size and achievement is non-linear such that the positive relationship plateaus after a certain school size is reached (Giambona & Porcur, 2018; Antoniou et al., 2024). One review of 57 empirical studies found that the effects of school size varied by who was served in the school (e.g. mostly disadvantaged students) and varied by age group (e.g. increases in size in secondary schools compared to elementary schools) (Leithwood and Jantzi, 2009). School size continues to be of interest in studies of school characteristics and academic achievement, but a clear answer remains elusive. School climate health is measured at the school level, so it is prudent to understand the possible relationship with a variable that is so commonly used to understand school characteristics.

Proportion students ready for school. TIMSS data has supported studies of the relationship between school readiness (ability to do early numeracy and literacy skills) and achievement. Within the eight highest-performing countries in TIMSS 2019, one recent study found that while both early literacy and numeracy were predictive of mathematics achievement, literacy skills were more highly predictive (Chang, 2023). School readiness is studied internationally and has benefitted from numerous large-scale, longitudinal studies that demonstrate the impact of early numeracy, literacy, and emotional development when starting school. Several studies in the US have followed students starting around or before age 4. One study of 33,717 students who had been assessed for school readiness at age 4 found positive relationships between school readiness in preschool and GPA, high-stakes assessment performance, and lower suspension rates in Grade 5 (Ricciardi et al., 2021). In another study, parents of 551 pairs of twins were asked to report literacy and numeracy skills in the children at kindergarten and found a positive correlation (r=0.47) with mathematics at age 8 (Valiente et al.,

2021). An earlier study conducted a latent class analysis to define clusters of readiness among a sample of 544 US children from 54 months (~4.5 years old) to predict academic and emotional skills in Grade 5 (Sabol & Pianta, 2012). The results offered more complexity as a result of the profile-based approach to the definition of school readiness. Those students who struggled with foundational literacy and numeracy at 54 months continued that pattern at the later checkpoint. However, readiness in cognitive skills like attention did not predict difficulties at 5th grade. Having limited relational skills at 54 months was found to be predictive of low achievement at 5th grade. Furthermore, the authors found that school readiness was predictive of mathematics achievement "above contextual and demographic factors through fifth grade" which the authors point out is contrary to literature that has found links between parent factors and achievement (Sabol & Pianta, 2012, p.295).

Studies have also taken a school-level approach. Among 425 middle schools in the US, a study found that the proportion of students not ready for school was negatively related to achievement as was the proportion of disadvantaged students (Etim et al., 2022). This relationship is highlighted not to target the students but to demonstrate the school-wide impact of not adequately supporting student needs. The relationship between school readiness and wellbeing is somewhat less explored. One study of 3907 Australian children based on teacher evaluations of school readiness and four facets of wellbeing in Grade 6 found that early emotional development was predictive of wellbeing but early cognitive evaluations were only predictive of sadness and worry (Gregory et al., 2021).

Proportion of students disadvantaged. Extant literature explores the numerous interactions of socioeconomic status and educational and societal outcomes. Within studies referenced in this section, most of them also contend with the advantages in academic

advancement of coming from high-earning families or living in well-resourced neighborhoods. TIMSS collects data in this arena from multiple perspectives and via varying estimations such as parent's educational attainment and the number of books in the home. For this study, the school-level variable was used, but has an issue of relativity in that the principal evaluates for themselves what disadvantaged might mean in giving the proportion of students in their school that are 'disadvantaged'. However, it is worthwhile to understand what predictive role having many disadvantaged students might have in school climate health cluster assignments. Recent studies suggest a positive relationship between socioeconomic status and wellbeing. In other words, having more wealth and advantage gives a greater likelihood of wellbeing. One study of 61,759 students in Grades 4 to 9 in Australia found that parents' educational attainment was predictive of student wellbeing and predictive of greater access to instruments for improving wellbeing like extracurricular activities (Kennewell et al., 2022).

There is empirical support for what we know to be true observationally – socioeconomic status alone is hugely influential in the life one has the opportunity to lead. Operating in the specific social networks of schools, the adults in the system can work to improve the environment that students experience. One study of 1550 Austrian students in disadvantaged schools found that positive teacher emotions and attitudes greatly contributed to overall student wellbeing at school (Holzer et al., 2024). This study will not define an overarching theory of the network of forces operating on disadvantaged students and their schools but is interested in knowing what school climate health clusters they mostly reside in.

Proportion speaking language of test. Multilingualism is an enriching skill linked to cultural heritage and practices that can nevertheless impact how students interact with their school community. Students of different language backgrounds than their schooling peers are a

heterogeneous group that can have different racial, immigration, and economic histories that can differentially impact their schooling. Numerous studies of multilingual students have found they experience school climate differently than their peers and can struggle with feeling included at school. Even in studies mentioned earlier, multilingual students are identified as groups for whom there are significantly different outcomes than the sample as a whole. However, studies find that even though sense of school belonging can be lower for these students (e.g. Juvonen et al., 2019), the influence and intentional intervention of teachers and school community members can compensate (van der Wildt, 2017). One study of school readiness and growth trajectories in 9216 refugee and immigrant children in Canada found that within this group, English Language Learners (ELLs) were more likely to struggle (Gagne et al., 2021). The study followed students who were identified in kindergarten as needing literacy and numeracy assistance and mapped their achievement trajectories between grades 4 and 10. The authors point out that immigrant and refugee students can be uniquely motivated to achieve compared to their peers. Levels of readiness in kindergarten were predictive of later trends defined by 'Thriving', 'Catching Up,' or 'Falling Behind'. The study found interactions in trajectory group membership based on socioeconomic status and ELL status – both were more likely to be in groups characterized as "Below Average" and "Declining". The study results are interesting on their own but is also an important reminder that heterogeneity exists even within groups that are defined as being alike.

Chapter Summary

Wellbeing is a complex summation of internal emotions and perceptions of context. International concern for children's wellbeing has been absorbed into educational endeavors through global shifts in what education should accomplish. Evaluating the extent of wellbeing requires an operationalizable framework and benchmarks. While numerous exist in response to

the former, there is minimal theorization or methodology to support the latter. The contention is that there is no isolatable measure or understanding of school belonging at the student level without instantaneously capturing the conditions of the system. Existing initiatives do not adequately address the qualities of the internal world of students internationally, nor the environments that might sustain them. Furthermore, operationalizing these frameworks has not pursued methods of simultaneous measurement to support these advanced frameworks. The literature on wellbeing and school climate supports the focus on relationships, environment, and stakeholder satisfaction. These three domains inform the first stage of selecting factors from the Happy School Framework that align with school climate health. The Happy Schools framework is primed for empirical validation through the realization of school climate health.

CHAPTER 3: ANALYSIS PLAN AND METHODOLOGY

The purpose of this chapter is to summarize the methodological approach taken in this dissertation. This includes outlining the four main objectives of the methodology, the analysis plan for the multilevel latent class analysis, and two secondary analyses conducted with school cluster membership. The study has four major phases:

A. Establish a framework of *school climate health* that can be measured by scales included in the TIMSS 2023 Context Questionnaires
B. Find latent classes of schools internationally based on multilevel latent class analysis (Vermunt, 2003) of the indicators identified in part A.
C. Describe the groups of schools that form the latent classes based on shared school characteristics and how these characteristics differ across classes. E.g.: Where, in geographical terms, are schools of each latent class mostly found?
D. Explore the relationship between academic achievement and different categorizations of healthy school climate

Part A of the research plan is detailed in the next section based on the selection of scales from the TIMSS 2023 Context Questionnaires supporting the definition of *school climate health*. Having isolated the importance of relationships, school environment, and stakeholder satisfaction, indicators are extracted from the UNESCO Happy Schools framework based on these key constructs. The application of the methods required in Parts B and C is supported by past applications of multilevel latent class analysis in education literature to identify profiles of school climate. Part D explores the predictive power of select demographic variables and variability in the relationships between academic attainment for students across school climate health clusters.

Adapting the Happy Schools Framework (Phase A)

Defining a gradient of quality from the outset is crucial to claim a satisfactory system state (Davidson, 2013). This research is interested in pursuing a novel approach to use ILSA data to support evaluations of school quality as understood to encompass a comprehensive view of system-level wellbeing sustained in schools. A working theory is needed to support empirical validation of how the numerous interwoven constructs of student wellbeing and school climate are operating in schools internationally. Hence, adapting the Happy School framework for this research relied on selecting components that relate to relationships, environment, and stakeholder perceptions. This established a subsample of the Happy Schools framework supported by existing empirical research for evaluating wellbeing and school climate.

This portion of the methodology to isolate aspects of the Happy Schools Framework that can be theoretically related to the empirically supported literature. The literature review illuminated three shared priorities across the well-being and school climate literature that could inform evaluative decisions: relationships, school environment, and stakeholder satisfaction. The first step of adapting the criteria from the Happy Schools Framework for analysis is to classify the criteria defined in that framework that relate to these three dimensions.

Isolating Relationships, Environments, and Stakeholder Satisfaction

The criteria defined in the Happy Schools framework are emergent themes drawn from the qualitative questionnaire conducted as part of its development. The first step in translating these criteria into the basis of school climate health was to draft brief definitions of these criteria based on the more detailed explanations given in the reporting of the Happy Schools project (UNESCO Bangkok, 2020). The contents here are specifically intended to assist the operationalization of empirical research of the Happy Schools in the context of the Grade 4

TIMSS audience. The Happy School criteria have been adapted to numerous other contexts and even re-organized for expansion to other cultures as encouraged by UNESCO and Happy Schools proponents. Thus, the present research is just one lens through which to view what is valuable in the project. These adapted definitions drawn from the Happy Schools Framework are included for reference in Table 3.1.

Table 3.1

Simplified Definitions of Happy Schools Criteria

People	Process	Place
Friendships and Relationships in the School Community • Student relationships with classmates and teachers; relationship between school management and also getting parents involved in the community **Positive Teacher Attitudes and Attributes** • Responsive to student and parent views on teacher performance – strictness, approach with students, in-class attitude **Respect for Diversity and Differences** • Eliminate favoritism and competition to ensure equitable distribution of attention and care to students. Teaching and curricular implications. **Positive Collaborative Values and Practice** • Sustain school environment supportive of empathy, kindness, tolerance, equal treatment **Teacher Working Conditions and Well-being** • Teacher recognition; being supported by salary, respect, acknowledgement; teacher motivation/sense of accomplishment **Teacher Skills and Competences** • Opportunities for professional development – how to make teaching more fun and innovative	**Reasonable and Fair Workload** • Heavily focused on the extent of high-stake assessments; focus on non-cognitive traits and supporting development of non-academic domains of learning. **Teamwork and Collaborative Spirit** • Student sense of belonging; student-student relationships supported through teaching practices **Fun and Engaging Teaching and Learning Approaches** • Student or Teacher perception of extent of enjoyable lessons **Learner Freedom, Creativity, and Engagement** • Extent that students feel free to make mistakes, free to express themselves. Teacher practices that support student communication are advised **Sense of Achievement and Accomplishment** • Extent students feel recognized and encouraged by teachers, parents, school community. Teacher/parent/school interventions include prizes, positive feedback and public recognition **Extracurricular Activities and School Events** • Extent students are offered non-academic after-school activities; school events outside school hours to increase student sense of school belonging **Learning as a Team Between Students and Teachers** • Positive collaborative friendly relationships between teachers and students **Useful, Relevant, and Engaging Learning Content** • Extent that curricula are responsive to contemporary and relevant issues; interdisciplinary teaching **Mental Well-Being and Stress Management** • Develop physical and psychological capacities; in-house psychologist; reflective practices	**Warm and Friendly Learning Environment** • Positive relationships with principals – colorful visual displays **Secure Environment Free from Bullying** • Environment that sustains bullying resulting from difference – all solutions are school-level **Open and Green Learning and Playing Spaces** • good infrastructure, sanitation facilities and ICT equipment; conducive to learning and socialization; contact with natural world **School Vision and Leadership** • Wider school management is committed to and communicates singular vision for school; promotes sense of belonging/achievement in students **Positive Discipline** • Extent that school policies and teacher practices are fair; want to motivate students rather than compare and punish; support students' emotional regulation **Good Health, Sanitation, and Nutrition** • Clean, healthy school environment; good nutrition **Democratic School Management** • Extent that stakeholders are heard in management considerations; students, parents, teachers have a voice

Considering a distilled definition was particularly helpful in cases where the label for the criterion was not wholly representative of later elaboration or guidance given elsewhere in the Happy Schools report. Two illustrative examples of this tension are "Open and Green Learning and Playing Spaces" and "Useful, Relevant, and Engaging Learning Content". Each label has good potential at face value to be included in the school climate health definition. "Open and Green Learning and Playing Spaces" evokes visions of children having positive educational environments for self-expression, and "Useful, Relevant, and Engaging Learning Content" seems to target the relationships between students and teachers that are at the heart of the quality relationships in school. However, the definitions and recommendations given in the Happy Schools Framework complicate this straightforward interpretation. Excerpts are included in Table 3.2 for each of these criteria. "Open and Green Learning and Playing Spaces" was defined in response to the need for *clean and sanitary* learning environments, with access to both educational technology and outdoor space at school; "Useful, Relevant, and Engaging Learning Content" was defined not to relate lessons to the everyday experience of students, but to ensure that what students learn is preparing them for adult life (UNESCO, 2020)

Table 3.2

Selected Examples of Happy Schools Criteria Definitions and Recommendations

	Open and Green Learning and Playing Spaces	**Useful, Relevant, and Engaging Learning Content**
Criteria Definition	"good infrastructure, sanitation facilities, and ICT equipment. In particular, respondents stressed the importance of green outdoor spaces for learning and playing, so that learners can spend time outside the classroom and connect with nature."	"At the national level, this requires that curricula be designed to reflect contemporary and relevant issues"
Report Recommendations	"Establish a school garden… Make use of outdoor spaces for learning and playing"	"using interdisciplinary classes to make learning more applicable to daily life and, in particular, to better prepare students for the world of work."

Following the initial definition exercise, Happy School criteria were considered for inclusion if they targeted the quality of the environment, relationships, or stakeholders i.e., were sufficiently responsive to school climate or well-being literature. The criterion label and its description were considered in unison to make an ultimate judgment. Returning to the "Green Spaces" and "Useful Content" examples, the descriptions did not match the purpose of school climate health in either case. "Green Spaces" targets the qualities of school facilities and physical spaces that are not universally accessible. As defined in Happy Schools, "Useful Content" is most related to high-level curricular decisions that may be out of school control depending on the local context. This research does not need to argue the benefits of outdoor recreation or high-quality curricula but must acknowledge that these are unfair criteria to apply at the school level as if they are equally available to all schools and necessary for high-quality school environments. The condition of physical school space is out of the control of students and school staff, as can be curricular decisions in many contexts. School climate health is interested in the *actions* that school community members can *implement* to support wellbeing and academic success.

The four criteria excluded (see Table 3.3) were Extracurricular Activities and School Events; Good Health, Sanitation, and Nutrition; Open and Green Learning and Playing Spaces; and Useful, Relevant, and Engaging Learning Content. Considering the more elaborate definitions, these categories are not suitable for defining school climate health, and removing them did not degrade the cohesiveness of the selected set of Happy School Criteria.

Table 3.3

Happy Schools Criteria Excluded from the First Round

Excluded Criteria	
Extracurricular Activities and School Events	Extent students are offered non-academic after-school activities; and school events outside school hours to increase student sense of school belonging
Good Health, Sanitation, and Nutrition	Clean, healthy school environment; good nutrition
Open and Green Learning and Playing Spaces	Good infrastructure, sanitation facilities, and ICT equipment; conducive to learning and socialization; contact with the natural world
Useful, Relevant, and Engaging Learning Content	Extent that curricula are responsive to contemporary and relevant issues; interdisciplinary teaching

The following phase was to match the relevant criteria with the available TIMSS 2023 Context Questionnaires by selecting scales that could act as indicators for the subset of criteria obtained to this point. This phase was a key step in operationalizing the framework and the literature but also presented a limitation in relying on the already finalized questionnaires. Table 3.4 includes the result of this process – each of the 22 criteria is defined, sorted into its most applicable category, and mapped to both a TIMSS 2023 and TIMS 2019 context questionnaire item or scale

Nonetheless, most criteria could be represented at least partially by a scale or group of items in the dataset. Four criteria that did not have a good representative in the TIMSS 2023 context questionnaires are displayed with grey boxes in Table 3.4. The results of this selection process include 14 of the 22 Happy Schools criteria and 22 scales or item groups from the TIMSS 2023 context questionnaires.

Table 3.4

Summary of Selected Criteria, Definitions, and Related TIMSS Scales

Environment		**T23**
Learner Freedom, Creativity, and Engagement	Develop physical and psychological capacities; in-house psychologist; reflective practices	G13 (T), G9(S)
Mental Well-Being and Stress Management	Environment that sustains bullying resulting from difference – all solutions are school-level	
Positive Collaborative Values and Practice	Extent that students feel free to make mistakes, free to express themselves. Teacher practices that support student communication are advised	G6 (T)
Reasonable and Fair Workload	Heavily focused on the extent of high-stake assessments; focus on non-cognitive traits and supporting development of non-academic domains of learning.	
School Vision and Leadership	Positive relationships with principals – colorful visual displays	12, 11
Secure Environment Free from Bullying	Sustain school environment supportive of empathy, kindness, tolerance, and equal treatment	G14 (S); G7 (T)
Warm and Friendly Learning Environment	Wider school management is committed to and communicates singular vision for school; promotes sense of belonging/achievement in students	G13 (S)
Relationships		
Friendships and Relationships in the School Community	Extent that school policies and teacher practices are fair; want to motivate students rather than compare and punish; support students' emotional regulation	14
Learning as a Team Between Students and Teachers	Student relationships with classmates and teachers; relationship between school management and also getting parents involved in the community	G12 (T) MS3, MS8
Positive Discipline	Eliminate favoritism and competition to ensure equitable distribution of attention and care to students. Teaching and curricular implications.	
Respect for Diversity and Differences	Positive collaborative friendly relationships between teachers and students	
Teamwork and Collaborative Spirit	Student sense of belonging; student-student relationships supported through teaching practices	MS4, MS9
Stakeholder		
Democratic School Management	Extent students feel recognized and encouraged by teachers, parents, school community. Teacher/parent/school interventions include prizes, positive feedback and public recognition	8(H)
Fun and Engaging Teaching and Learning Approaches	Extent that stakeholders are heard in management considerations; students, parents, teachers have a voice	
Positive Teacher Attitudes and Attributes	Opportunities for professional development – how to make teaching more fun and innovative	15
Sense of Achievement and Accomplishment	Responsive to student and parent view on teacher performance – strictness, approach with students, in-class attitude	MS5, MS10
Teacher Skills and Competences	Student or Teacher perception of extent of enjoyable lessons	MS2, MS7
Teacher Working Conditions and Well-being	Teacher recognition; being supported by salary, respect, acknowledgment; teacher motivation/sense of accomplishment	G8(T), G9 (T)

TIMSS Context Questionnaire Overview

The purpose of the TIMSS context questionnaires is to gather information internationally on factors of educational environments that impact mathematics and science achievement as measured by TIMSS (Mullis & Fishbein, 2020). Although not part of the trend measure, many of the constructs measured by the questionnaires have been consistently administered over numerous cycles of TIMSS assessments - evidence of their longstanding importance to the international education community At the fourth grade, questionnaires are distributed to sampled students, their teacher(s), school principal, and parents or guardians. The context questionnaire scales are designed to each target one construct that has been shown to relate to mathematics and science achievement. Scales vary in length and format but include most commonly between 8 and 12 items using Likert-type items with four response options. TIMSS develops these questionnaires each cycle to provide context and evidence to support policies and practices internationally.

Scores estimated based on these scales are used to establish two or three cut-points that define meaningfully different response patterns across the set of items. TIMSS employs Item Response Theory (IRT) modeling, specifically the Rasch partial credit model (Masters, 1982), to construct scales from context questionnaire items (Yin & Fishbein, 2020). For instance, the Student Bullying scale included in Figure 3.1 below is separated into three regions based on scale scores to define categories describing the frequency of feeling bullied: Never or Almost Never, About Monthly, and About Weekly (Mullis et al., 2020). These scores allow for ease of interpretation with the table of results also presented as part of the International Results publication where the percent of students falling within each region is reported by country. This exhibit also reports the average achievement for students in each category based on the same

scale cut scores. Inferences are possible regarding the relationship between student bullying and mathematics and science achievement (as in Figure 3.2).

Figure 3.1

Student Bullying Items Administered in TIMSS 2019

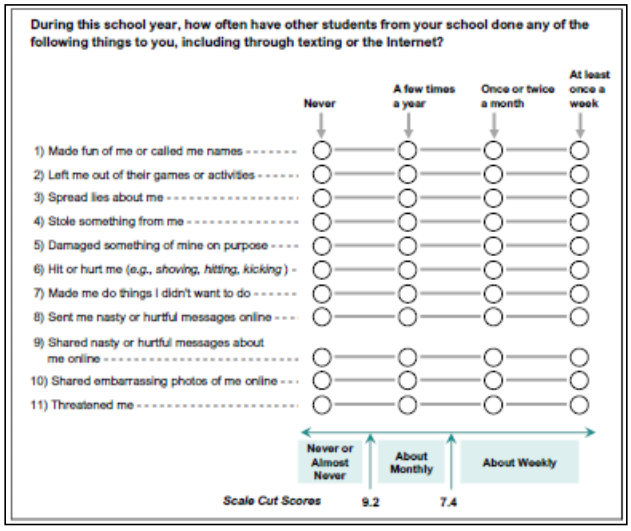
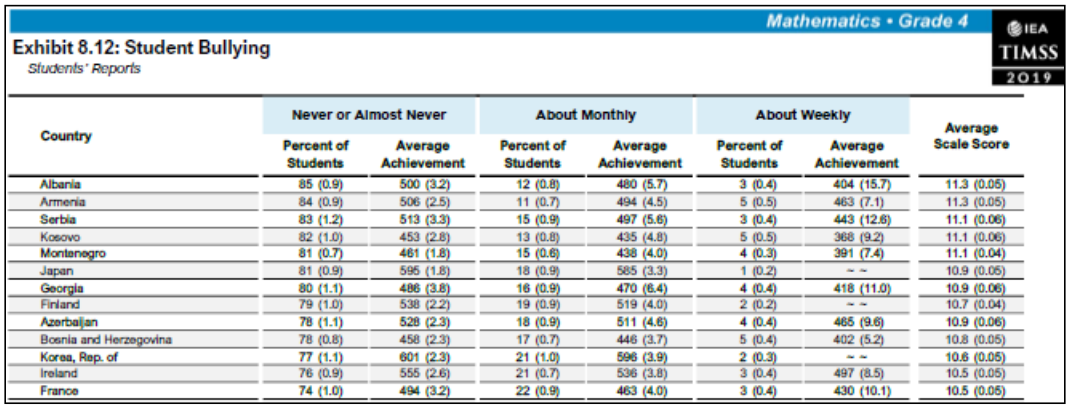


Figure 3.2

Student Bullying as Reported in TIMSS 2019 International Results Report



Mathematics • Grade 4

Exhibit 8.12: Student Bullying

Students' Reports

Country	Never or Almost Never		About Monthly		About Weekly		Average Scale Score
	Percent of Students	Average Achievement	Percent of Students	Average Achievement	Percent of Students	Average Achievement	
Albania	85 (0.9)	500 (3.2)	12 (0.8)	480 (5.7)	3 (0.4)	404 (15.7)	11.3 (0.05)
Armenia	84 (0.9)	506 (2.5)	11 (0.7)	494 (4.5)	5 (0.5)	463 (7.1)	11.3 (0.05)
Serbia	83 (1.2)	513 (3.3)	15 (0.9)	497 (5.6)	3 (0.4)	443 (12.6)	11.1 (0.06)
Kosovo	82 (1.0)	453 (2.8)	13 (0.8)	435 (4.8)	5 (0.5)	368 (9.2)	11.1 (0.06)
Montenegro	81 (0.7)	461 (1.8)	15 (0.6)	438 (4.0)	4 (0.3)	391 (7.4)	11.1 (0.04)
Japan	81 (0.9)	595 (1.8)	18 (0.9)	585 (3.3)	1 (0.2)	~ ~	10.9 (0.05)
Georgia	80 (1.1)	486 (3.8)	16 (0.9)	470 (6.4)	4 (0.4)	418 (11.0)	10.9 (0.06)
Finland	79 (1.0)	538 (2.2)	19 (0.9)	519 (4.0)	2 (0.2)	~ ~	10.7 (0.04)
Azerbaijan	78 (1.1)	528 (2.3)	18 (0.9)	511 (4.6)	4 (0.4)	465 (9.6)	10.9 (0.06)
Bosnia and Herzegovina	78 (0.8)	458 (2.3)	17 (0.7)	446 (3.7)	5 (0.4)	402 (5.2)	10.8 (0.05)
Korea, Rep. of	77 (1.1)	601 (2.3)	21 (1.0)	596 (3.9)	2 (0.3)	~ ~	10.6 (0.05)
Ireland	76 (0.9)	555 (2.6)	21 (0.7)	536 (3.8)	3 (0.4)	497 (8.5)	10.5 (0.05)
France	74 (1.0)	494 (3.2)	22 (0.9)	463 (4.0)	3 (0.4)	430 (10.1)	10.5 (0.05)

For questionnaire data from other stakeholders, the interpretation is slightly less direct. Although the same procedures are followed for parent, teacher, and principal data, inferences are still at the student level. For example, the Parents' Perceptions of Their Child's School scale (see Figure 3.3) similarly has three categories identified by two cut scores (Mullis et al., 2020).

However, the percentages reported are still through the students, i.e., a high percentage of students in a country have parents who are satisfied with their school. Thus, parents' perceptions can still be related directly to TIMSS achievement (see Figure 3.4). These examples are given to familiarize the reader with the format of the TIMSS questionnaires, but also to emphasize that all context questionnaires pertain to the scholastic experience of students sampled to participate in TIMSS.

Figure 3.3

Parents' Perceptions of Their Child's School Items Administered in TIMSS 2019

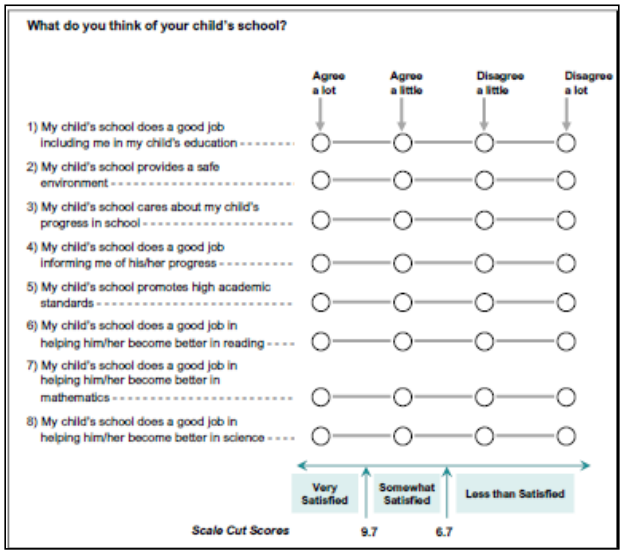


Figure 3.4

Parents' Perceptions of Their Child's School as in TIMSS 2019 International Report

Mathematics • Grade 4

Exhibit 7.7: Parents' Perceptions of Their Child's School

Students' Results based on Parents' Reports

IEA TIMSS 2019

Country	Very Satisfied		Somewhat Satisfied		Less than Satisfied		Average Scale Score
	Percent of Students	Average Achievement	Percent of Students	Average Achievement	Percent of Students	Average Achievement	
Albania	90 (0.8)	497 (3.4)	9 (0.7)	484 (8.6)	2 (0.4)	~ ~	11.5 (0.05)
Armenia	88 (1.0)	500 (2.6)	11 (0.9)	486 (5.5)	1 (0.1)	~ ~	11.5 (0.05)
Kazakhstan	87 (1.0)	513 (2.6)	12 (0.9)	514 (5.0)	1 (0.2)	~ ~	11.5 (0.05)
Malta r	85 (0.7)	516 (1.7)	14 (0.7)	515 (3.5)	1 (0.2)	~ ~	11.2 (0.03)
Kosovo	84 (0.8)	448 (3.0)	14 (0.6)	440 (5.2)	2 (0.3)	~ ~	11.1 (0.04)
Saudi Arabia	80 (0.9)	404 (3.7)	17 (0.7)	390 (6.1)	3 (0.3)	393 (10.8)	11.0 (0.05)
Northern Ireland s	79 (1.2)	590 (3.9)	19 (1.1)	566 (5.5)	2 (0.3)	~ ~	11.0 (0.06)
Georgia	79 (1.1)	482 (3.8)	19 (1.0)	490 (4.8)	2 (0.2)	~ ~	10.9 (0.05)
North Macedonia	79 (1.1)	473 (5.9)	18 (0.9)	482 (7.7)	3 (0.5)	473 (10.9)	10.8 (0.05)
Azerbaijan	78 (1.1)	524 (2.8)	19 (1.0)	514 (4.3)	3 (0.3)	514 (8.0)	10.9 (0.06)
Bulgaria	78 (1.4)	513 (4.0)	19 (1.1)	529 (7.0)	4 (0.4)	526 (15.3)	10.9 (0.07)
Montenegro	77 (0.8)	455 (2.2)	20 (0.8)	454 (3.4)	3 (0.3)	455 (10.3)	10.9 (0.03)

Analysis Plan (Phases B, C, and D)

Using student responses on school climate questionnaires to identify and explore school profiles employing latent class analysis has precedent. Galvez-Nieto et al. (2020) conducted a multilevel latent cluster analysis using response data from nearly 3000 students – four different surveys asked about their perceptions of the school, their peers, their teachers, and their community. Their final model selection described 6 student clusters and two classes of schools, those with overall poor school environment ratings and those with positive school environment ratings (Galvez-Nieto, 2020, p. 90). This study pursues similar methodologies but aims to find patterns across schools and groups of countries at the higher level. Taking the exploration of multi-faceted conceptualizations of school climate to another level, this project includes school perception data from different *categories* of individuals with different roles in schools. Where Galvez-Nieto and colleagues combined student perspectives on different facets of school climate, this study also includes teacher, parent, and principal perceptions of facets of school climate.

Data and Sample

The data of interest are responses to 22 scales administered as part of the TIMSS 2023 Grade 4 Context Questionnaires. These 22 scales consist of one scale (8 items) from parent responses, four scales (41 items) from principal responses, 11 scales (81 items) from student responses, and six scales (49 items) from teacher responses. A total of 179 items have been selected and grouped to form the three major theme groups:

School Environment consists of 76 items from eight scales;
Stakeholder Satisfaction has 56 items from eight scales, and;
Positive Relationships has 47 items from six scales.

The scales chosen as the indicator variables are presented in Table 3.5 along with the theme and Happy Schools Framework Criteria they align with, collectively.

Table 3.5

Overview of Scales and Items Selected and Happy Schools Criteria

Relationships	Environments	Stakeholder Satisfaction
ATBG12 – *Engaging Teacher Practices* ASBGDLM – *Disruptive Lessons, Mathematics* ASBGDLS – *Disruptive Lessons, Science* ASBGICM – *Instructional Clarity, Mathematics* ASBGICS – *Instructional Clarity, Science* ACBGDAS – *Principal's Report, Disorder*	ATBGEAS – *Teacher's Report, Academic Emphasis* ATBGSLN – *Teacher' Report, Students not Ready for Instruction* ATBGSOS – *Teacher's Report, School Discipline* ASBGSSB – *Student Sense of Belonging* ASBG09 – *Students Arrive Hungry/Tired* ASBGSB – *Student Bullying* ACBGEAS – *Principal's Report, Academic Emphasis* ACBGMRS – *Principal's Report, Limited Resources*	ATBGTJS – *Teacher Job Satisfaction* ATBG09 – *Non-Instruction Responsibilities* ASBGSLM – *Students Like Mathematics* ASBGSLS – *Students Like Science* ASBGSCM – *Students Confident Mathematics* ASBGSCS – *Students Confident Science* ASBHPSP – *Parents Satisfied with School* ACBG15 – *Principal's Report, Teacher Absenteeism*
Friendships and Relationships in the School Community Learning as a Team Between Students and Teachers Teamwork and Collaborative Spirit	Learner Freedom, Creativity, and Engagement Positive Collaborative Values and Practice School Vision and Leadership Secure Environment Free from Bullying Warm and Friendly Learning Environment	Democratic School Management Positive Teacher Attitudes and Attributes Sense of Achievement and Accomplishment Teacher Skills and Competences Teacher Working Conditions and Well-being

The international versions of the TIMSS 2023 Context Questionnaires will be made available with the publication of the International Results report and International Database. All but three items use a four-response Likert scale, with some individual items using wording requiring reverse coding. Of these three items, two have five response choices relating to the extent to which each item is emphasized in the school, and the remaining one has three response options. Scores on scales and summed scores across items were aggregated to the school level then recoded into categorical variables – this process is reported later as part of Chapter 4.

While the final database for TIMSS 2023 is not yet available, the participating countries that are included in the analysis are summarized in Table 3.6.

Table 3.6

TIMSS 2023 Grade 4 Countries in Analysis

Countries			
Albania	Denmark	Kosovo	Romania
Armenia	England	Kuwait	Saudi Arabia
Australia	Finland	Latvia	Serbia
Azerbaijan	France	Lithuania	Singapore
Bahrain	Georgia	Macao SAR	Slovak Republic
Belgium (Flemish)	Germany	Montenegro	Slovenia
Belgium (French)	Hong Kong SAR	Morocco	South Africa
Bosnia & Herzegovina	Hungary	Netherlands	Spain
Brazil	Iran	New Zealand	Sweden
Bulgaria	Ireland	North Macedonia	Turkiye
Canada	Italy	Norway	United Arab Emirates
Chile	Japan	Oman	United States
Chinese Taipei	Jordan	Poland	Uzbekistan
Cyprus	Kazakhstan	Portugal	
Czech Republic	Korea, Rep. of	Qatar	

Multilevel Latent Class Analysis

Multilevel latent class analysis (MLCA) (Vermunt, 2003) is an appropriate choice for the present research purpose to extract school climate health profiles at the school level. Latent class analyses are statistical models that extract group similarities by identifying subgroups that share

response patterns on indicator variables for the model (Bakk et al., 2013). The multilevel specification is the most appropriate considering the way TIMSS samples classrooms in schools and schools within countries. Specifically, multilevel extensions of latent structure models are more appropriate in the case of complex sampling as is the case with TIMSS data (von Davier, 2007). Schools in the same country are likely to provide more similar responses than schools across countries (Bakk, di Mari, Oser, and Kuha, 2022). Therefore, multi-level or hierarchical models are required to properly take this clustering of schools in countries into account. MLCA has been employed in exploring classroom and school-level climate groupings based on self-reported questionnaire data (Galvez-Nieto et al., 2020; Allison et al., 2016).

Clustering is a fundamental design feature of sampling in TIMSS. Any analytic model must take this into account to enable appropriate analysis – the data fails the assumption of independence of observations within clusters preventing a simpler analysis (Henry and Muthen, 2010). Nationally representative samples for countries participating in TIMSS assessments were collected by stratified random sampling of schools from the population of schools teaching students at the target grade in that country. The probability of sampling is based on stratification variables identified by the country, with classrooms within chosen schools selected with equal probability (LaRoche, Joncas, and Foy, 2020). Respondents to all instruments administered during TIMSS (i.e. achievement and questionnaire) are selected for participation as a result of their relationships to one of the sampled classrooms within a chosen school. Teachers, principals, and parents are administered Context Questionnaires based only on their relationship to a selected classroom.

In the planned analysis, 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. 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 model is based on a probabilistic model that allows the log-odds of Level 1 class membership priors (class sizes) to vary across countries. The classes 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 classes that are identified in the 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. Secondary analysis expands this definition by explaining any defining characteristics among schools within a school climate health latent cluster. For instance, if one group has the highest scores of Stakeholder Satisfaction and also has the highest level of a socio-economic variable, both would be included in the description of the class even though the latter is not a measure of school quality.

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 possible values are 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 (Vermunt, 2003). Distinctions are evident in the logit equations, below first assuming independence across countries *J*,

$$P(X_{ij} = t) = \frac{\exp(\gamma_t)}{\sum_{r=1}^{s_k} \exp(\gamma_r)}$$

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

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

$$P(X_{ij} = t) = \frac{\exp(\gamma_{tj})}{\sum_{r=1}^{T} \exp(\gamma_{rj})}$$

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

Non-parametric model. Two main approaches exist in specifying an MLCA: the parametric model and the non-parametric model. In pursuing the non-parametric model over the parametric model, the chosen Level 1 model becomes the basis of predicting latent class membership at Level 2. The parametric model requires specification of how many classes exist at Level 2, while the non-parametric defines them based on the Level 1 model (Vermunt, 2003; Asparouhov & Muthen, 2014). The non-parametric model relaxes normality requirements and has a lower computational burden in estimating the model. Pursuing the nonparametric approach to multilevel latent class analysis in the response data allows latent class sizes to be defined at the country level. 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):

- Step 1: Simple Latent Class Analysis
  - Find the number of latent classes in the data without any clustering 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 based on BIC, classification error, and entropy R^2
- Step 3: Predictors for Level 1 class membership with multinomial logistic regression

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. Each school is assigned to exactly one cluster extracted by the model. Classification error is present in both methods and are frequently used (Bakk, Takle, and Vermunt, 2013). This research employed the modal method as the

number of indicators specified and the multilevel specification resulted in singular response patterns.

Secondary Analyses

Executing Step 3 of the MLCA analysis plan has two major possibilities – including covariates into the clustering model directly (e.g. Yamaguchi, 2000 or Muthen 2004), or undertaking external analyses to predict cluster membership based on variation in those covariates (e.g. Chan and Goldthorpe, 2008) or class membership is a predictor (e.g. Vermunt, 2003). Previous research using MLCA in school climate profiles has included covariates of interest directly into the latent model after initial modeling steps are concluded. Covariate effects are given by odds-ratio comparison of being a member in a particular class compared to the parallel value of the covariate (Gohari et al., 2020). For example, Galvez-Nieto et al., (2020) analyzed school climate latent class clusters in a small student sample based on responses to four questionnaires sampled from 32 schools. Level 1 was the student, and Level 2 was the school. Covariates at Level 2 included type of school, while at Level 1 variables for age, sex, and family typology were included as predictors. One study of college student attitudes included self-report variables as covariates such as career interest and personality traits like neuroticism or openness (Mutz & Daniel, 2013).

However, the present study chose to pursue two analyses external to the MLCA estimation i.e. after cluster assignment was established based on the final model specification. Many examples of applied research choose to conduct secondary analyses outside of the latent class estimation via the so-called three-step approach (Bakk, Tekle, and Vermunt, 2013). By conducting a separate analysis, the exploratory cluster definition can be conducted without considering the influence of covariates, and the possibility for confounding variables is

introduced as, for instance, an avenue for further exploration rather than integral to the findings of international school climate health patterns. This approach is theoretically resonant with this study where demographic or structural variables are of interest yet not defining school climate health. Thus, only perceptions and attitudes of individuals participating in the school community were included in the MLCA, but further definition of school conditions within school climate clusters is possible. Similarly, achievement is not a defining variable for school climate health but the extent to which cluster membership is predictive of achievement is of interest to this research.

Thus, multinomial logistic regression was employed to evaluate the predictive power of demographic school factors that may influence school classification of school climate health. Finding a high odds-ratio would indicate the presence of a confounding factor in the similarities defining a group cluster. The relationship between group assignment and TIMSS 2023 Grade 4 Mathematics achievement was investigated via analysis of variance with school average achievement as the outcome variable and school climate grouping added alongside known predictors of achievement. These analyses could provide information to support claims that the qualities of the school climate are predictive of achievement over and above the predictive contribution of other known predictors.

A note of caution is prudent that the parameter estimates tend to be deflated as a function of the classification error resulting from the probabilistic MLCA model (Bakk & Kuha, 2018). Since the parameter values are interpreted here only to indicate possible interactions from variables other than the target construct, no adjustment is made to correct this. For further reading, see Bakk, Tekle, and Vermunt, 2013.

Phases B and C – Conducting MLCA Analysis

Specification of a model with both school clusters and country classes adequately responds to the research questions of the dissertation overall. Summaries of preparing the data, specifying the models, and interpreting the output are presented. A preliminary interpretation of the results demonstrates how typologies of schools on school climate health are informed and how countries can be classified based on the proportion of school clusters within.

Preparing the Data

In preparing the data for the analysis LatentGOLD, the final data file must be formatted with one entry per school per country and all variables aggregated to the school level. For all data files, this required cleaning and reorganizing in R. The R code is provided in Appendix B. For scales in all four files the basic steps were as follows:

1. Concatenate IDCOUNTRY and IDSCHOOL to create UNIQUE_SCHOOL variable.
2. For Teacher, Student, and Parent data files, calculate the school mean of scale scores with appropriate sampling weights (TOTWGT for students/parents; TCHWGT for teachers) via Means function in SPSS
    a. Five selected item groups were not subjected to scaling, so scale scores were not available. A required intermediate step was to calculate the score by individual across those items. This was achieved by adding the item responses per individual.
3. All data files combined based on the UNIQUE_SCHOOL variable. SCHWGT is the school-level sampling weight variable and was included for each school.

Conducting the Analysis in LatentGold

Past applications of MLCA have been modeled for up to a dozen clusters at both individual and group levels (Fagginger Auer et al., 2016). Multilevel approaches to latent class analysis are meaningful beyond just their appropriateness to the nested data in ILSA applications because groups at both levels can be specified and interpreted separately as well as integrated.

Phases B, C, and D are addressed by or rely on the results from the MLCA. Phase B comprised executing the MLCA and choosing the best model by school and country clusters following the three steps to modeling from Bakk, di Mari, Oser, and Kuha, (2022). LatentGold estimates the likelihood of group membership based on different response patterns on the model indicators. Numerous fit statistics are calculated to judge overall model fit in the software to choose the best model to move forward with interpreting. Phase C follows directly from the output of the final MLCA model; LatentGold reports the proportion of all respondents identified as being members of the clusters of the final model. Thus, having defined the model it was possible to identify the school characteristics and country similarities based on the group assignments from the final model chosen.

Model Definition Overview. To specify a model for analysis, the variables included as indicators have to be defined, the number of clusters specified, and various technical specifications that are required by the software are used.. The response variables from the TIMSS 2023 Grade 4 Context Questionnaire were included as indicators for all models using all country data. At the start of the modeling process, a range of clusters can be estimated at once with the indicator variables chosen. However, one must specify an exact cluster number to get predicted class membership output for analysis outside of LatentGOLD which is relevant to Phase D of the analysis plan. In setting up the model to be estimated, the multilevel extension also supports complex sampling designs by indicating a stratum or applying weights. To estimate the multilevel extension of the latent class analysis, a Group ID variable (Country) is specified. However, to enact this model, at least 2 GClass clusters must be specified (Vermunt and Magidson, 2005; Vermunt and Magidson, 2016).

Further specification is possible in the remaining setup tabs to adjust, for example, the treatment of missing data. The default specification is to exclude cases with missingness. However, the TIMSS dataset has incomplete records and represents aggregated data from multiple respondent types. Thus, all models were specified to include all cases regardless of missingness.

Output and Interpretation. Modeling is an exercise that requires different iterations of estimates 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.

Different indicators may be more or less influential in deciding on final models. The first step as outlined by Bakk, di Mari, Oser, and Kuha, (2022) is to judge the fit of the models specifying different numbers of clusters at the school level (Level 1). In this study the Bayesian Information Criteria (BIC, [Schwarz, 1978]), classification error, and an entropy pseudo-R^2 were compared with each model estimation (Vermunt & Magdison, 2016). The BIC is widely used in numerous modeling applications beyond just latent class analyses (Lee and Stankov, 2023; Asparouhov & Muthen, 2014) as a conservative measure considering more cautiously the number of degrees of freedom and parameters compared to other information criteria that can be generated. When comparing any two models, there is no target value or indicator of sufficiency, but instead the BIC can indicate improvement in fit or signal a turning point after which adding more clusters does not improve the model.

Classification error was also considered as model complexity increased. Defined as the estimated proportion of classification error based on the posterior membership probabilities, this measure can indicate how successfully the model separates cases into classes based on the chosen indicators. Across the data, schools are assigned to clusters based on the maximum membership probability, such that the proportion of classification error essentially takes the weighted average of the inverse of a school's maximum posterior membership probability taken over all schools in the data (Vermunt & Magdison, 2016). Modeling decisions can also be supported by changes to entropy R^2. With a range from 0 to 1, entropy R^2 gives a summary of the predictive qualities of the model to correctly assign cases to clusters – values over 0.8 are considered a good classification result (Aspourhov & Muthen, 2014).

In addition to fit statistics, too many clusters could lead to problems with interpretability as the proportion of cases that populate a cluster may be too small. Even with a large sample size, clusters with less than 10% of cases can be difficult to support with substantive interpretation (Lee & Stankov, 2023). LatentGOLD gives several proportional descriptive statistics regarding membership in the estimated classes.

Estimation with group-level (country) classes follows a similar approach. LatentGOLD assigns countries to clusters based on the extent a group of countries has a similar proportion of each cluster based on the schools assigned to the extracted clusters. Revealing latent classes of countries supports the hypotheses of this research by supporting typologies that are emergent from the data rather than assigned to the results based on geography or politics. The steps to finalize the model are to determine the number of group-level clusters and then re-estimate the Level-1 clusters under that restriction.

Example Output. LatentGOLD provides several tables of results on cluster composition to support the interpretation of the classification of both levels of the model. For example, consider the results from a modeling exercise resulting in 3 country classes and 4 school clusters. The proportion of the sample in each of the three country-level classes is reported, along with the proportion of each GClass belonging to the four Level-1 clusters (See Figure 3.5). Across the four Level-1 clusters, 36.43% of schools belong to Cluster 1, 23.78% to Cluster 2, 20.77% to Cluster 3, and 19.02 to Cluster 2.

Figure 3.5

LatentGOLD Output – Cluster proportions

	GClass			
	1	2	3	
	0.5303	0.3030	0.1667	
	Cluster			
GClass	1	2	3	4
1	0.5443	0.0933	0.2690	0.0934
2	0.1527	0.3910	0.0694	0.3870
3	0.1761	0.4191	0.2644	0.1404

Along with the mean values per variable per country-level Group and school-level Class, for each categorical variable included in the model, LatentGOLD output also reports the likelihood that a member in a cluster would give that response (see Figure 3.6).

Figure 3.6

LatentGOLD Output– Membership proportion by response category

	Cluster1	Cluster2	Cluster3	Cluster4
Overall	0.3643	0.2378	0.2077	0.1902
Indicators				
ACBG15				
1 - 12	0.4427	0.0043	0.0000	0.5530
13 - 16	0.6259	0.3702	0.0039	0.0000
17 - 40	0.0000	0.3677	0.6323	0.0000
.	0.4212	0.1731	0.2713	0.1343

Investigation of the grouping outside of LatentGOLD is readily facilitated; for any model estimated, group assignments per school are exported into SPSS files. This file supports secondary analyses with demographic variables at the school level (i.e. size, proportion speaking the language of test, or socioeconomic status) not included as indicator variables in the cluster model. Among the avenues for further investigation with the SPSS export are country-level assignments to groups such that country similarities can be interpreted per group. Similarly, the proportion of clusters within each country is readily summarized in SPSS. An example from three countries (Albania, Armenia, and Azerbaijan) is also included in Table 3.7 to show the results of the posterior classification. These tables contribute to meaningfully interpretating where different school climate health compositions reside globally.

Table 3.7

Proportion of Cluster Membership in Three Example Countries

Country	Cluster	Schools	Percent
Albania	2	64	38.3
	4	103	61.7
Armenia	1	2	1.3
	2	70	46.7
	3	1	.7
	4	77	51.3
Azerbaijan	1	16	8.2
	2	101	52.1
	3	3	1.5
	4	74	38.1

Caution is warranted to define the cluster too specifically, as variation in aggregated scores across countries could reflect cultural response style differences as much as variation in school climate health (Harzing, 2006). However, following the deliberate selection of variables and judicious modeling process justifies looking for commonalities in the groupings of countries that the model reveals.

Chapter Summary

The analysis plan for this research has four major stages as outlined in this chapter. Phase A comprises the alignment exercise to operationalize the Happy Schools Framework criteria with TIMSS 2023 Grade 4 Context Questionnaire variables. Phases B-D represent the main statistical analyses of this research. The multilevel extension of latent class analysis is employed to extract school- and country-level grouping based on probabilistic estimations of membership derived from response patterns. Demographic variables are included in two secondary analyses to investigate the possible influence of confounding variables in the latent class specification and the relationship to mathematics achievement. An exploratory approach to model selection is summarized along with an overview of the model definition and LatentGOLD output.

CHAPTER 4 – RESULTS

The MLCA included twenty-two indicator variables that were each aggregated to the school level. These comprised four from the principal questionnaires, six from the teacher questionnaires, eleven from the student questionnaires, and one from the parent questionnaires. The dataset is sourced from 58 countries participating in TIMSS 2023, with responses from 359,053 students, 22,731 teachers, and 12,014 school principals. 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. For items that required summation to create individual level scores, a result of 0 was indicative of total non-response on that group of items. For TIMSS questionnaire items, each response is associated with a score of 1 or more depending on the number of response categories. Thus, a score of 0 is only possible from a group of items if they are entirely left blank. These cases were recoded as missing after the summation was completed. There was otherwise no direct intervention into missingness and the default treatment of the means functions in SPSS were conducted.

One special note is warranted for countries not administering the home questionnaire. Some countries elect not to administer the Home Questionnaire to parents – for these countries no values for ASBHPSP are included, but this does not require they be excluded from the analysis.

Model Preparation

Descriptive statistics for the MLCA indicator variables are included in Tables 4.1- 4.3 before any coding was implemented. The indicators are organized by respondent – Principal, Teacher, and Student/Home. Both median and mean statistics and associated standard errors are provided for the 449,859 schools internationally represented by the weighted data. When interpreting the scale scores, it is prudent to refer to the Rasch scale cutscores that define three regions per scale – these are included as part of the TIMSS 2023 International report (forthcoming).

Table 4.1

Descriptive Statistics of MLCA Indicator Variables (School Weight Applied) – Principal

	ACBG15	ACBGMRS	ACBGEAS	ACBGDAS
Valid	414908	418431	417107	414960
Missing	34951	31427	32752	34899
Mean	3.05	10.09	9.79	9.85
(s.e)	.00	.00	.00	.00
(s.d)	1.56	2.30	2.17	1.88
Median	2.00	10.00	9.52	10.08

Table 4.2

Descriptive Statistics of MLCA Indicator Variables (School Weight Applied) – Teacher

	ATBGLSN	ATBGEAS	ATBGSOS	ATBGTJS	ATBG09	ATBG12
Valid	424587	427194	426787	426284	425702	426945
Missing	25271	22665	23071	23575	24157	22914
Mean	9.74	9.87	10.19	10.03	17.63	10.85
(s.e)	.00	.00	.00	.00	.01	.00
Median	9.70	9.72	10.35	10.09	17.51	10.50
(s.d)	1.95	2.17	1.97	1.88	4.83	3.17

Table 4.3

Descriptive Statistics of MLCA Indicator Variables (School Weight Applied) – Student/Home

	ASBGSSB	ASBGSB	ASBGSLM	ASBGICM	ASBGDML	ASBGSCM	ASBGSLS	ASBGICS	ASBGDSL	ASBGSCS	ASBHPSP	ASBG09
Valid	447027	446851	446335	446435	446010	445542	445616	446045	445418	445512	345724	447041
Missing	2831	3008	3523	3424	3849	4316	4242	3814	4441	4347	104135	2818
Mean	10.05	9.89	10.17	10.00	9.89	9.94	10.19	10.02	9.89	10.03	10.19	4.65
(s.e)	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Median	9.98	9.86	10.14	10.01	9.73	9.88	10.20	10.07	9.73	9.96	10.34	4.64
(s.d)	.98	.94	1.01	.94	1.13	.78	.92	.91	1.15	.88	1.11	.96

Each of the 22 indicator variables required recodes for the MLCA analysis in LatentGOLD. 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. LatentGOLD can also automatically create ordinal variables with user-specified categories, but there is less flexibility in labeling the resulting categories. Thus, SPSS 29 was used to calculate the cutpoints and execute the recoding. 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 Table 4.4-4.6. In the interpretation, these categories are interpreted as "Low", "Medium", and "High", respectively.

Table 4.4

Recode for MLCA Indicator Variables (School Weight Applied) - Principal

	ACBG15	ACBGDAS	ACBGEAS	ACBGMRS
Cutpoint 1	2.00	9.15	8.86	9.13
Cutpoint 2	3.00	10.45	10.50	10.74

Table 4.5

Recode for MLCA Indicator Variables (School Weight Applied) - Teacher

	ATBGLSN	ATBGEAS	ATBGSOS	ATBGTJS	ATBG09	ATBG12
Cutpoint 1	8.86	8.99	9.32	9.09	15.50	9.00
Cutpoint 2	10.34	10.60	11.10	11.21	19.51	12.00

Table 4.6

Recode for MLCA Indicator Variables (School Weight Applied)) –Student/Home

	ASBHPSP	ASBGSSB	ASBGSB	ASBGSLM	ASBGICM	ASBGDML	ASBGSCM	ASBGSLS	ASBGICS	ASBGDSL	ASBGSCS	ASBG09
Cutpoint 1	9.82	9.57	9.49	9.73	9.68	9.35	9.56	9.77	9.68	9.31	9.61	4.29
Cutpoint 2	10.78	10.37	10.23	10.59	10.37	10.19	10.21	10.61	10.42	10.23	10.33	5.00

Modeling and Results of MLCA

Modeling is an iterative process leading to a final decision on a model that best fits the data and research purpose. The exercise balances statistical estimates and evaluative judgements. For the MLCA in this study, LatentGOLD supports these overall judgments by producing a single table with different statistics for every model run on a dataset. Models were run and evaluated in three stages as described in Chapter 3.

Level 1 Estimation

The first comparison is to judge the fit of the models specifying different numbers of clusters at the school level (Level 1). All twenty-two indicator variables were included as ordinal variables with three categories each following the recode described earlier. The weighting variable appropriate for school-level analysis (schwgt) was included (for further information see Vermunt & Magidson, 2016). All cases were included regardless of missingness. A range of clusters was specified from 1-10, and the resulting fit statistics are summarized in Table 4.7. LatenGOLD calculates numerous fit statistics to support modeling decisions – based on the

priorities of the analysis, different indicators may be more or less influential in deciding on final models

Table 4.7

Summary of Model Fit Statistics - Level-1 Models

		LL	BIC(LL)	Npar	Max. BVR	Class.Err.	Entropy R^2
Model1	1-Cluster	-10331524.20	20663621.12	44	290381.54	0.000	1.000
Model2	2-Cluster	-9828531.46	19657935.04	67	228954.18	0.045	0.844
Model3	3-Cluster	-9637630.23	19276431.96	90	88064.54	0.070	0.839
Model4	4-Cluster	-9526489.01	19054448.90	113	95283.09	0.097	0.823
Model5	5-Cluster	-9440571.45	18882913.17	136	81346.60	0.108	0.824
Model6	6-Cluster	-9380264.13	18762597.92	159	64916.81	0.123	0.821
Model7	7-Cluster	-9340411.59	18683192.21	182	63567.67	0.133	0.818
Model8	8-Cluster	-9299857.78	18602383.99	205	68476.27	0.130	0.830
Model9	9-Cluster	-9265556.65	18534081.11	228	61453.92	0.145	0.823
Model10	10-Cluster	-9240709.21	18484685.60	251	36992.63	0.158	0.816

Turning to these first ten models, the BIC decreases with each additional cluster which suggests that model fit improves with increased complexity. However, the other fit statistics provide balance in deciding what model might be best. Classification error is lowest in the simplest model (2-cluster; 0.045), and the entropy R^2 is highest here too (2-cluster; 0.844). However, all of the Level-1 models (Models 1- 10) are above the entropy 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 (0.045 to 0.070) and again between 3-clusters and 4-clusters (0.070 to 0.097). The entropy R^2 is steady across the 4-, 5-, and 6-cluster models (0.0823, 0.0824, 0.0821, respectively).

In addition to problems with interpretability with more complex cluster models, the proportion of cases that populate a cluster may be too small. Even with a large sample size, clusters with less than 10% of cases can be difficult to support with substantive interpretation

(Lee & Stankov, 2023). In this case, cluster proportions to trending towards 10% after the 5-cluster model, where two clusters have only 14% and 13% of cases. With priority to interpretability, and considering these indicators, the best candidates are the 3- and 4-cluster models.

Because the final stage of the MLCA modeling process involves specifying simpler models with the chosen G-Class, the 4-cluster model is chosen.

Level-2 Estimation

Estimation with group-level classes follows a similar approach. LatentGOLD assigns countries to clusters based on the extent two countries have similar proportions of cluster membership across their schools. Fixing the number of Level-1 clusters to 4, models with 2-5 group classes (GClass) were specified. Revealing latent classes of countries supports the hypotheses of this research by supporting typologies that emerge from the data rather than being assigned based on an priori conditions such as geography or politics. The results of the models from these steps are summarized in Table 4.8.

Table 4.8

Summary of Model Fit Statistics - Level-2 Models

		LL	BIC(LL)	Npar	Max. BVR	Class.Err.	Entropy R^2
Model21	4-Cluster 1-GClass	-9526489.02	19054448.90	113	95283.09	0.097	0.823
Model22	4-Cluster 2-GClass	-9440187.61	18881898.17	117	103498.99	0.088	0.839
Model23	4-Cluster 3-GClass	-9381828.09	18765231.20	121	109074.91	0.084	0.850
Model24	4-Cluster 4-GClass	-9376293.09	18754213.27	125	109540.88	0.083	0.851
Model25	4-Cluster 5-GClass	-9367562.30	18736803.75	129	110057.18	0.082	0.852

The BIC decreases with each additional country class specified and the classification error increases between the 2-, 3-, and 4- GClass models. 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.

While the 3- and 4-GClass models are similar in terms of model fit statistics, investigating the modal assignment of countries in each model reveals a clearer choice. For the 4-Gclass model, the fourth country class only has eight countries which is close to 10% 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 distraction more than information. To avoid the groupings being too small, we conclude that three country classes are the best fit for the data.

Re-estimating the school-level models at Level 1with the 3-GClass specification is the last stage of the modeling process. The resulting model fit statistics are included in Table 4.9, and based on these alone the best-fitting model remains elusive. 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.

Table 4.9

Summary of Model Fit Statistics – Level 1 Models with Level-2 Fixed

		LL	BIC(LL)	Npar	Max. BVR	Class.Err.	Entropy R^2
Model36	1-Cluster 3-GClass	-10331524.20	20663647.16	46	290381.54	0.000	1.000
Model37	2-Cluster 3-GClass	-9759214.50	19519353.18	71	239499.23	0.037	0.866
Model38	3-Cluster 3-GClass	-9517955.25	19037160.11	96	94753.04	0.059	0.865
Model23	4-Cluster 3-GClass	-9381828.091	18765231.20	121	109074.91	0.084	0.850

Visual representations from the LatentGold output were investigated to aid in determining if the additional class (4 compared to 3) may hold value in interpretation. Table 4.10 contains three profile graphs produced by LatentGold showing the mean score per cluster over

the model variables. These are included to illustrate that the four-cluster model essentially shares the basic composition of the three in the three-cluster model.

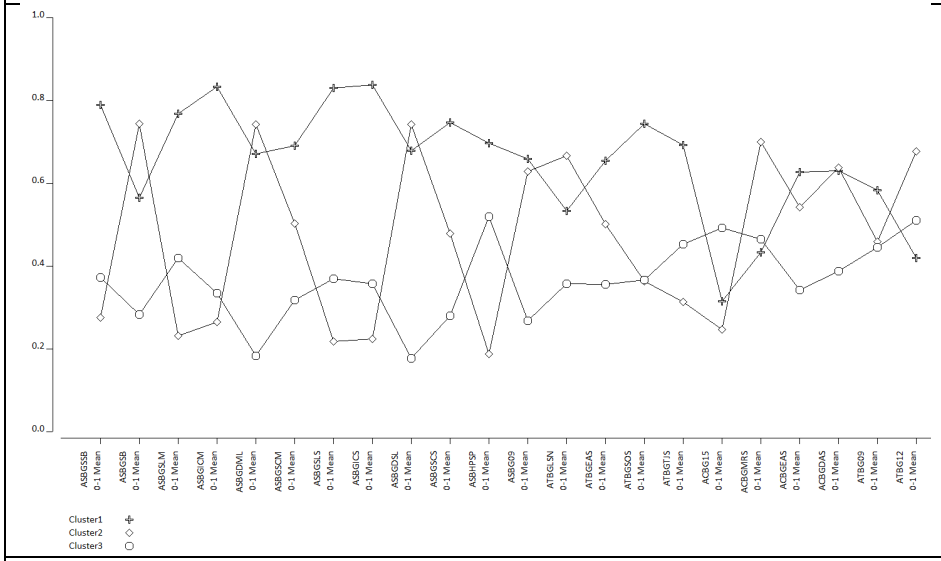
The additional distribution does not impact the shared characteristics between the 3- and 4- cluster figures. Interpreting the 4-cluster graphs at once is less straightforward than the apparent 'high' 'middle' 'low' pattern of the 3-cluster graphs. However, isolating the 'high' cluster in the 4-cluster 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. This differential in experience is an interesting result for this study given that it describes a phenomenon directly referenced in the literature.

Thus, the 4-Cluster 3-GClass model is selected as the final model for school climate health.

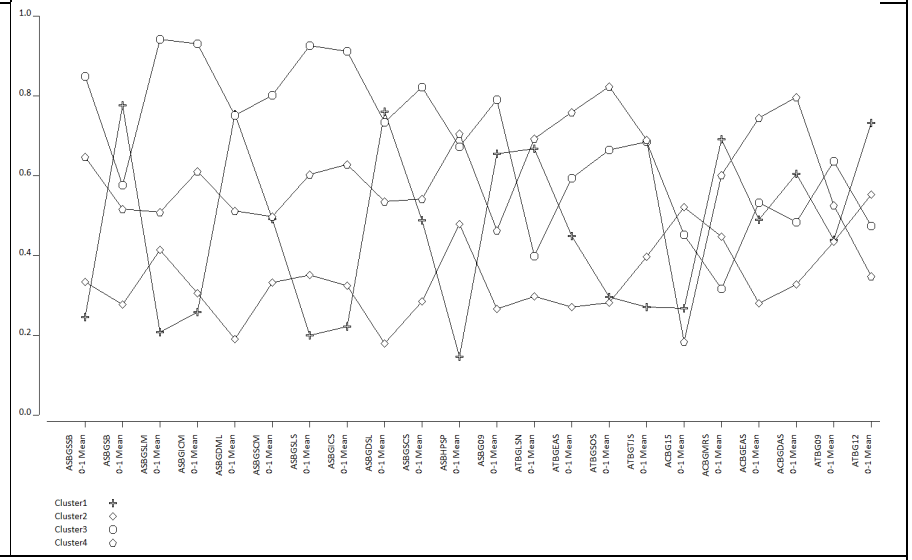
Table 4.10

Comparison of Graphs Resulting from 3-Cluster and 4-Cluster Solution

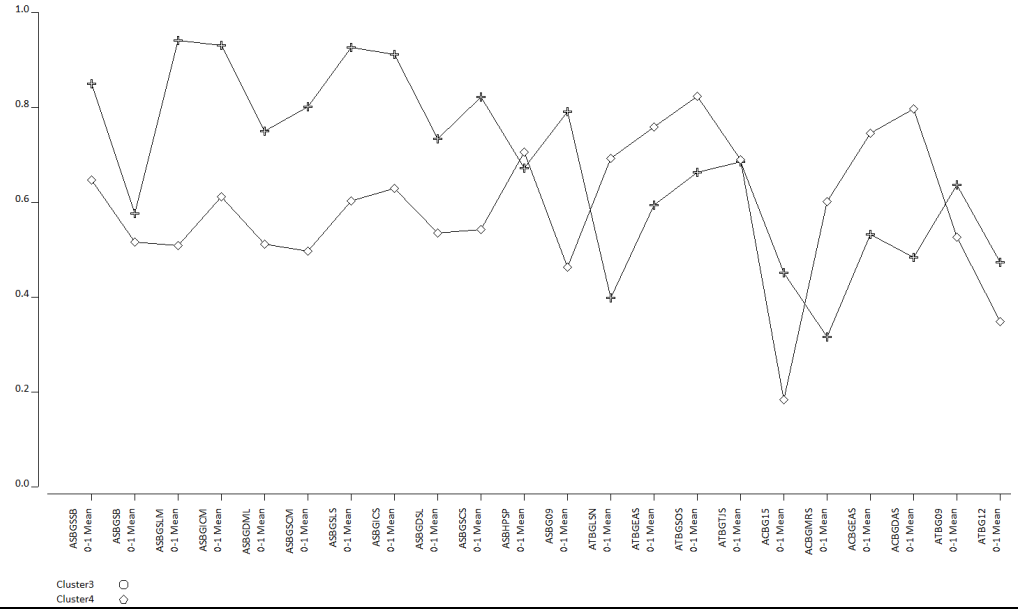
3-Cluster



4-Cluster



Isolation of additional cluster against 'high' Cluster



Cluster Composition

The results of the MLCA can be reported at each level and then integrated. First, the Level-1 4-Cluster model results will be summarized at the school level, followed by a description of the 3-Class Level-2 groupings. To understand what characterizes these Level-2 groupings, the results at both levels are combined to report the proportions of each Level-1 cluster that are in the country groups.

The 4-Cluster, 3-Class model that was chosen in the modeling process resulted in four groups of relatively equal proportion. The clusters are not ordered in the course of the analysis but will be referred to as Cluster 1-4 until names and characteristics are further elaborated later in this section. Based on the posterior membership probabilities, schools were assigned to the four clusters such that 27% are in Cluster 1, 27% in Cluster 2, 25% in Cluster 3, and 21% in Cluster 4 (as in Table 4.11). The almost uniform distribution is not a consequence or characteristic of the MLCA – in other words, it is not the case that any 4-cluster solution would exhibit this sort of case distribution.

Table 4.11

Proportion of Cluster Composition (Modal Assignment)

	Cluster1	Cluster2	Cluster3	Cluster4
Cluster Size	0.275	0.267	0.253	0.206

Parameter estimates are given for each modeled cluster and indicator variable. However, the most relevant statistics associated with the parameter estimates are the Wald statistic, its p-value, and the R^2. A non-significant p-value would indicate that the indicator is not well-distinguished across clusters (see Vermunt & Magidson, 2016). Due to the large sample size in the TIMSS dataset, the statistical significance is an unhelpful indicator of which variables contributed most to the separation of school cases into clusters as is done in studies with smaller

samples. While the Wald statistic does still increase for variables contributing to greater separation, it should be considered in union with the R^2. This 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 indicator and the total variance (for further detail, see Vermunt & Magidson, 2006). These results are presented in descending order by R^2 in Table 4.12.

Table 4.12

MLCA Output - Parameters and Variance Statistics

	Cluster1	Cluster2	Cluster3	Cluster4	Wald	p-value	R^2
ASBGSLS	-1.48	-0.77	2.10	0.15	77567.75	5.3e-16842	0.46
ASBGICM	-1.23	-1.02	2.12	0.13	76527.35	4.4e-16616	0.44
ASBGICS	-1.33	-0.86	1.93	0.26	85418.72	8.5e-18547	0.44
ASBGSLM	-1.45	-0.57	2.25	-0.24	59334.83	7.8e-12883	0.43
ASBGSSB	-1.08	-0.70	1.38	0.40	83447.93	7.5e-18119	0.36
ASBGDSL	0.84	-1.51	0.72	-0.05	83156.87	1.2e-18055	0.35
ASBGDML	0.80	-1.44	0.78	-0.14	81474.43	2.6e-17690	0.34
ASBHPSP	-1.44	-0.01	0.66	0.79	58177.10	1.9e-12631	0.30
ATBGSOS	-0.70	-0.75	0.41	1.04	69315.23	5.1e-15050	0.28
ASBG09	0.36	-1.00	0.93	-0.29	65837.17	9.0e-14295	0.25
ASBGSCS	-0.19	-0.90	1.11	-0.02	57765.79	4.0e-12542	0.23
ASBGSB	0.88	-0.91	0.11	-0.08	57160.27	1.2e-12410	0.21
ATBGTJS	-0.81	-0.36	0.57	0.59	54702.91	4.8e-11877	0.20
ATBGEAS	-0.23	-0.88	0.24	0.87	55864.58	2.7e-12129	0.19
ASBGSCM	-0.16	-0.69	0.99	-0.14	45039.71	9.5e-9779	0.18
ATBGLSN	0.50	-0.72	-0.37	0.59	50615.27	1.9e-10989	0.17
ACBGDAS	0.11	-0.66	-0.22	0.76	45442.84	2.8e-9866	0.15
ACBGEAS	-0.07	-0.71	0.05	0.72	44632.73	2.3e-9690	0.14
ACBGMRS	0.55	-0.20	-0.61	0.26	29945.13	4.3e-6501	0.12
ATBG12	0.65	0.06	-0.17	-0.54	31325.41	8.4e-6801	0.11
ACBG15	-0.21	0.45	0.28	-0.52	28409.31	1.3e-6167	0.09
ATBG09	-0.21	-0.23	0.39	0.05	13132.19	2.2e-2850	0.04

Most of the indicator variables included in the model discriminated well between the clusters in the final cluster model. The student variables (aggregated to the school level) had the highest R^2 values, with teacher and school variables following with lower values. Low R^2 values,

as for variables ATBG09 with only 4% of variance explained, do not imply that it should 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.

The four most discriminating variables are school averages for four scales of student perceptions of classroom functioning and student attitudes towards mathematics and science. ASBGICM and ASBGICS relate to student perceptions of their teacher's clarity in instruction in mathematics classes and science classes, respectively. ASBGSLM and ASBGSLS relate to student reports of liking mathematics and science, respectively.

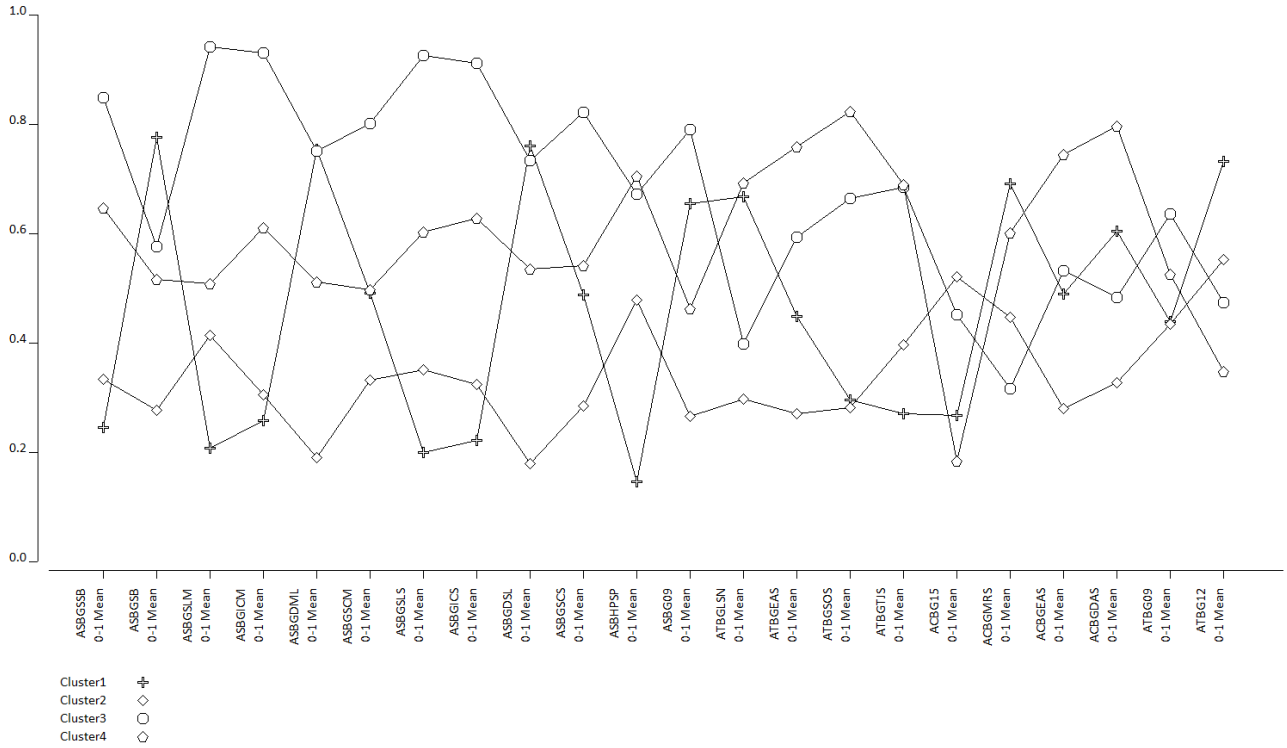
Following these, student sense of school belonging (ASBGSSB) and student perceptions of disorderly behavior in the classroom interrupting instruction in their mathematics and science lessons are the most discriminating (ASBGDML and ASBGDSL, respectively). The only home questionnaire variable also has a relatively high R^2 (ASBHSP, R^2= 0.3), representing the extent that parents are satisfied with their child's school.

The extent of the discrimination per variable is efficiently represented in the graph output from the model, shown in Figure 4.1. The y-axis represents the cluster mean, which LatentGold rescales to be between 0 and 1 for ordinal variables (Vermunt & Magidson, 2016). The visual spread corresponds with variables with varying discriminating ability. These highly discriminating indicators are good candidates for developing definitions based on characteristics that are distinct to each group. Most of the indicator variables included in the model discriminated well between the clusters in the final cluster model. For the most discriminating

variables (ASBGICM, ASBGICM, ASBGSLS, ASBGSLM), one can track the variation across clusters. Those variables with R^2 less than 0.1 did not demonstrate much separation across the clusters.

Figure 4.1

Cluster Separation Across Indicator Variables



The tables reporting proportions of each response category given by cluster supports interpretation of the highly discriminating variables that those which are less highly discriminating. Table 4.13 shows the proportional responses per response category for schools assigned to each cluster. For all variables, response categories resulting from the recode described earlier can be understood to be 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. Additionally, the descriptions follow the analysis in that they are reflective of school-level attributes. Following the TIMSS convention, high scores indicate the

desirable outcome, for instance, a high score on ASBGDML or ASBGDSL indicates *little disorder* in the classroom, not the abundance. The one exception is ATBG12 where a high score indicates rarely engaging in a given teacher practice.

For both ASBGICM and ASBGICS, Cluster 1 has a small proportion of schools with high instructional clarity (less than 0.1 responding in the High category). Looking to just the mathematics scale (ASBGICM), for schools in Cluster 1 the majority of responses are in the Low category (0.57). Cluster 2 schools have similar characteristics to Cluster 1, with roughly half in the low category and one-third in the Medium. Cluster 3 is the complement to these characteristics, with almost no schools (0.007) in the Low, and the outsize majority in the High category (0.87). The separation in these clusters is evident in these response patterns - the groupings defined by the MLCA have distinct characteristics. Finally, Cluster 4 has a relatively equal split between Medium and High categories and a small proportion in the Low. Values reported here are for ASBGICM, but the science scale version ASBGICS follows the same patterns (see Table 4.13).

Looking at ASBGSLS and ASBGSLM, schools in Cluster 1 once again have the highest proportion of responses in the Low category. For both mathematics and science, roughly 60% of schools in Cluster 1 have lower levels of students liking those subjects on average compared to the whole sample of schools. Cluster 3 is again the complement with over 80% of schools in that cluster in the High category, reporting high levels of liking mathematics and science, on average. Clusters 2 and 4 have a more complementary relationship than was observed in ASBGICM and ASBGICS. Schools in both clusters are roughly split between two response categories with about 40% in each. In Cluster 2, these are the Low and Medium categories, while in Cluster 4 it is the Medium and High categories.

Table 4.13

Response Percentages per Cluster by Respondent

Student Questionnaire Variables									
ASBGICM	Cluster 1	Cluster 2	Cluster 3	Cluster 4	ASBGDML	Cluster 1	Cluster 2	Cluster 3	Cluster 4
0	0.569	0.506	0.007	0.178	0	0.085	0.673	0.087	0.279
1	0.347	0.380	0.125	0.422	1	0.324	0.274	0.327	0.419
2	0.085	0.114	0.868	0.400	2	0.592	0.053	0.587	0.302
Mean	0.516	0.608	1.861	1.222	Mean	1.507	0.381	1.500	1.023
ASBGICS					ASBGDSL				
0	0.623	0.482	0.012	0.166	0	0.080	0.690	0.096	0.255
1	0.313	0.389	0.154	0.412	1	0.321	0.263	0.342	0.422
2	0.065	0.130	0.835	0.422	2	0.599	0.046	0.562	0.323
Mean	0.442	0.648	1.823	1.256	Mean	1.519	0.356	1.466	1.068
ASBGSLM					ASBG09				
0	0.645	0.380	0.006	0.279	0	0.168	0.577	0.073	0.346
1	0.294	0.415	0.107	0.427	1	0.355	0.314	0.274	0.384
2	0.061	0.205	0.887	0.294	2	0.477	0.109	0.654	0.271
Mean	0.415	0.826	1.882	1.015	Mean	1.308	0.532	1.581	0.925
ASBGSLS					ASBGSCS				
0	0.655	0.448	0.008	0.186	0	0.322	0.556	0.056	0.269
1	0.292	0.404	0.132	0.423	1	0.380	0.321	0.244	0.378
2	0.053	0.148	0.859	0.391	2	0.298	0.123	0.700	0.352
Mean	0.398	0.700	1.851	1.205	Mean	0.976	0.568	1.644	1.083
ASBGSSB					ASBGSCM				
0	0.598	0.481	0.037	0.165	0	0.326	0.502	0.072	0.320
1	0.315	0.370	0.228	0.379	1	0.367	0.333	0.255	0.367
2	0.087	0.149	0.735	0.457	2	0.307	0.164	0.673	0.313
Mean	0.489	0.667	1.698	1.292	Mean	0.981	0.662	1.601	0.993
ASBHPSP					ASBGSB				
0	0.745	0.323	0.149	0.124	0	0.084	0.568	0.240	0.297
1	0.219	0.398	0.359	0.341	1	0.279	0.313	0.369	0.375
2	0.037	0.280	0.493	0.535	2	0.637	0.119	0.392	0.328
Mean	0.292	0.957	1.344	1.411	Mean	1.553	0.552	1.152	1.031

Table 4.13

MLCA Output - Response Percentages per Cluster by Respondent (cont.)

School Questionnaire Variables				
ACBG15	Cluster 1	Cluster 2	Cluster 3	Cluster 4
0	0.657	0.391	0.462	0.753
1	0.151	0.175	0.175	0.128
2	0.192	0.433	0.363	0.120
Mean	0.535	1.042	0.901	0.367
ACBGMRS				
0	0.167	0.395	0.541	0.244
1	0.286	0.318	0.288	0.312
2	0.547	0.287	0.171	0.445
Mean	1.381	0.892	0.631	1.201
ACBGEAS				
0	0.366	0.596	0.324	0.137
1	0.289	0.248	0.288	0.238
2	0.345	0.156	0.388	0.625
Mean	0.980	0.560	1.064	1.488
ACBGDAS				
0	0.272	0.557	0.390	0.109
1	0.247	0.234	0.255	0.189
2	0.480	0.210	0.355	0.702
Mean	1.208	0.653	0.965	1.593

Teacher Questionnaire Variables				
ATBGLSN	Cluster 1	Cluster 2	Cluster 3	Cluster 4
0	0.170	0.548	0.430	0.150
1	0.325	0.310	0.345	0.315
2	0.505	0.142	0.225	0.535
Mean	1.335	0.594	0.795	1.385
ATBGEAS				
0	0.362	0.573	0.222	0.094
1	0.378	0.314	0.370	0.295
2	0.260	0.113	0.408	0.611
Mean	0.897	0.540	1.186	1.517
ATBGSOS				
0	0.575	0.592	0.200	0.078
1	0.260	0.255	0.274	0.200
2	0.165	0.154	0.526	0.723
Mean	0.590	0.562	1.326	1.645
ATBGTJS				
0	0.583	0.432	0.157	0.153
1	0.294	0.343	0.317	0.315
2	0.122	0.225	0.526	0.532
Mean	0.539	0.792	1.369	1.379
ATBG09				
0	0.394	0.399	0.205	0.306
1	0.334	0.333	0.319	0.337
2	0.272	0.267	0.477	0.357
Mean	0.878	0.868	1.272	1.051
ATBG12				
0	0.139	0.294	0.373	0.508
1	0.260	0.308	0.310	0.289
2	0.602	0.398	0.318	0.203
Mean	1.463	1.104	0.945	0.696

The proportions of responses given by cluster can be interpreted in union with the mean scale scores given before the recoding – these are summarized in Table 4.14. The patterns of high and low response percentages can be matched with the mean scores given in the table.

Table 4.14

Means on Indicator Variables by Cluster

	1		2		3		4	
	Mean	(s.e)	Mean	(s.e)	Mean	(s.e)	Mean	(s.e)
ASBGSSB	9.43	0.002	9.66	0.002	10.98	0.003	10.34	0.002
ASBGSB	10.48	0.002	9.40	0.002	10.13	0.004	9.92	0.003
ASBGSLM	9.45	0.003	9.94	0.002	11.34	0.002	10.15	0.003
ASBGICM	9.40	0.003	9.59	0.002	11.02	0.002	10.24	0.002
ASBGDML	10.47	0.003	9.13	0.002	10.67	0.004	9.86	0.003
ASBGSCM	9.90	0.003	9.63	0.001	10.57	0.003	9.92	0.002
ASBGSLS	9.48	0.002	9.87	0.002	11.19	0.002	10.41	0.002
ASBGICS	9.38	0.003	9.67	0.002	10.96	0.002	10.28	0.002
ASBGDSL	10.55	0.003	9.06	0.002	10.61	0.004	9.94	0.003
ASBGSCS	9.96	0.002	9.54	0.002	10.86	0.003	10.10	0.002
ASBHPSP	9.13	0.004	10.20	0.003	10.69	0.003	10.74	0.003
ASBG09	4.95	0.002	4.18	0.002	5.40	0.003	4.50	0.003
ATBGLSN	10.51	0.006	8.86	0.004	9.39	0.007	10.58	0.005
ATBGEAS	9.56	0.006	8.66	0.004	10.48	0.009	11.22	0.006
ATBGSOS	9.39	0.005	9.12	0.005	10.98	0.006	11.60	0.004
ATBGTJS	8.97	0.006	9.55	0.005	10.92	0.006	10.77	0.005
ATBG09	16.73	0.015	16.90	0.012	19.44	0.018	17.94	0.015
ATBG12	12.45	0.010	11.03	0.008	10.57	0.011	9.65	0.008
ACBG15	2.63	0.004	3.36	0.004	3.69	0.008	2.47	0.003
ACBGMRS	10.90	0.007	9.73	0.006	9.14	0.007	10.66	0.008
ACBGEAS	9.78	0.006	8.71	0.005	10.00	0.008	11.06	0.006
ACBGDAS	10.18	0.005	9.12	0.004	9.33	0.008	10.95	0.005

The least discriminating variables have differing patterns but share internal uniformity that illuminates further how so little variation is explained by the cluster assignment. ATBG09 asked teachers to report the extent they agreed with eight statements regarding possible stresses on their teaching including too many teaching hours, students, or administrative tasks. The

variable included in the MLCA was aggregated to the school level. In Clusters 1, 2, and 4, schools in that cluster respond about equally in each of the three response categories – roughly one-third belong to each of the High, Medium, and Low response categories. Cluster 3 stands out with roughly half of its schools in the High category, indicating that teachers in these schools, on average, are not experiencing the stresses on their time and position to the extent that the rest of the school sample is. For each of these indicator variables based on the average response given across clusters, it is clear to see how they did not contribute to distinguishing the clusters overall. However, there is still insight in some, such as in the case of Cluster 3 at the variable ATBG09, which stands out among the uniformity of the others on that variable.

ACBG15 asked principals to report the extent that late arrivals and absenteeism in teachers were a problem at their school. In Clusters 1 and 4, the great majority responded in the Low response category and Cluster 2 and 3 have nearly half of their schools in the same category. In clusters 1 and 4, there was also similarity across the other response categories across clusters as well – between 10% and 20% across the Medium and Low categories. Clusters 2 and 3, though, have a higher proportion of schools responding in the High category than the Low, suggesting that these teacher issues are more problematic for more schools in these clusters. While the variation accounted for by the clustering alone is not a high for ACBG15 as other variables, there is still valuable characteristic information to be extracted from the cluster separation.

Cluster Description Summaries

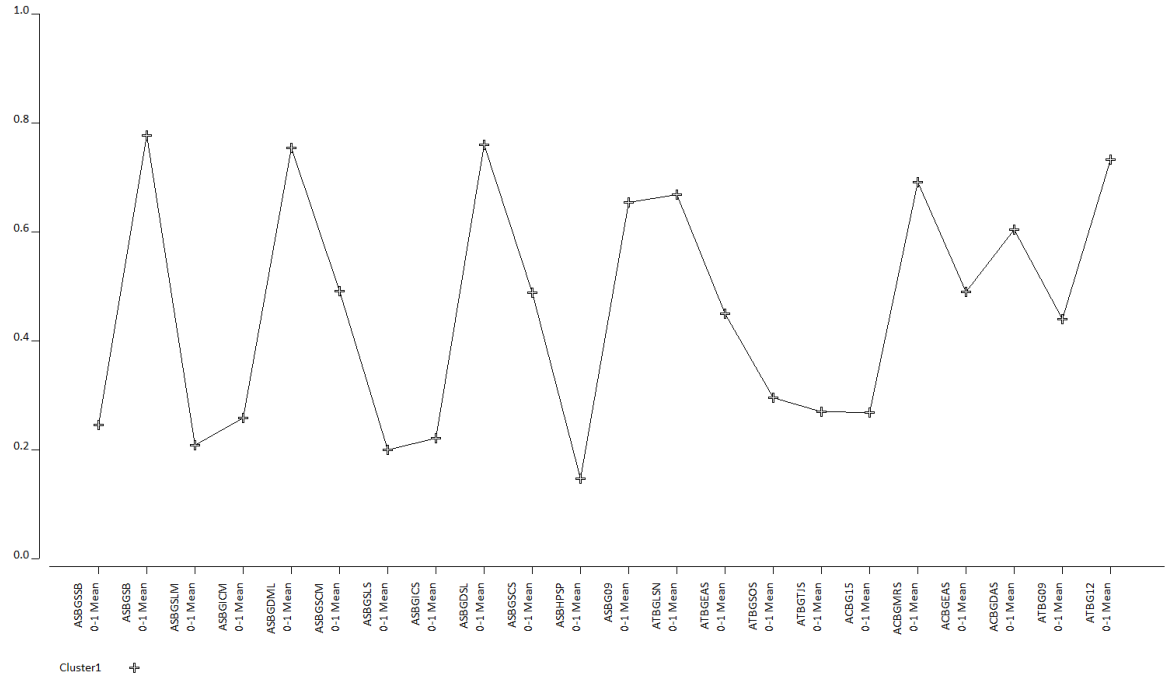
Cluster 1 – System Misalignment

Cluster 1 is comprised of school climate health contradictions. It has the highest mean on ASBGSB (student bullying), and ASBGDML/ASBGDSL (disorderly mathematics and science

classes, respectively) indicating that students are *rarely* bullied and experience *highly* orderly classrooms. Further, teachers report on average to be least limited by students not ready for instruction in their classes (ATBGLSN). Principals also report little disruption to instruction due to resource shortages (ACBGMRS), and few disciplinary issues at the school (ACBGDAS). However, schools in this cluster have the lowest of any cluster in average student measures including, sense of belonging (ASBGSSB), liking mathematics and science (ASBGSLM and ASBGSLS, respectively), and reports of instructional clarity in mathematics and science (ASBGICM and ASBGICS). Parents are least satisfied with their child's school (ASBHPSP), teachers are least satisfied with their jobs (ATBGTJS), and do not feel their school is aligned in discipline (ATBGSOS). Furthermore, teachers relate lessons to their students' lives least (ATBG12).

Figure 4.2

MLCA Indicator Means Plot – Cluster 1



Looking further into two of these somewhat anomalous patterns, it is helpful to review the cluster mean responses before the recode for the MLCA was conducted. The average score for the ATBGSOS score for teachers in schools in Cluster 1 was 9.39 – this is aligned with the middle, "Somewhat Safe and Orderly", scale region as reported in the TIMSS 2023 International report (forthcoming). This is helpful to understanding these results as it underlines that these schools do not necessarily have a problem, but that in relation to the other clusters they share a common sense of less order. Furthermore, this scale targets teacher perceptions of student respect and school disciplinary communication rather than the teachers feeling physical harm. The average score for the ASBGSSB score for students in schools in Cluster 1 was 9.43 which is within the middle scale category as reported in the TIMSS 2023 International report – Some Sense of Belonging.

Cluster 1 can be characterized as a group of schools where the priorities and values of the stakeholders involved are misaligned. Teachers and schools are well-resourced and have a calm, orderly school – yet there is little coherence in the school environment. Students feel lost in both their sense of belonging and in their class time, leading to their dislike of their subjects. Teachers and parents are both dissatisfied, and teachers don't feel they are respected.

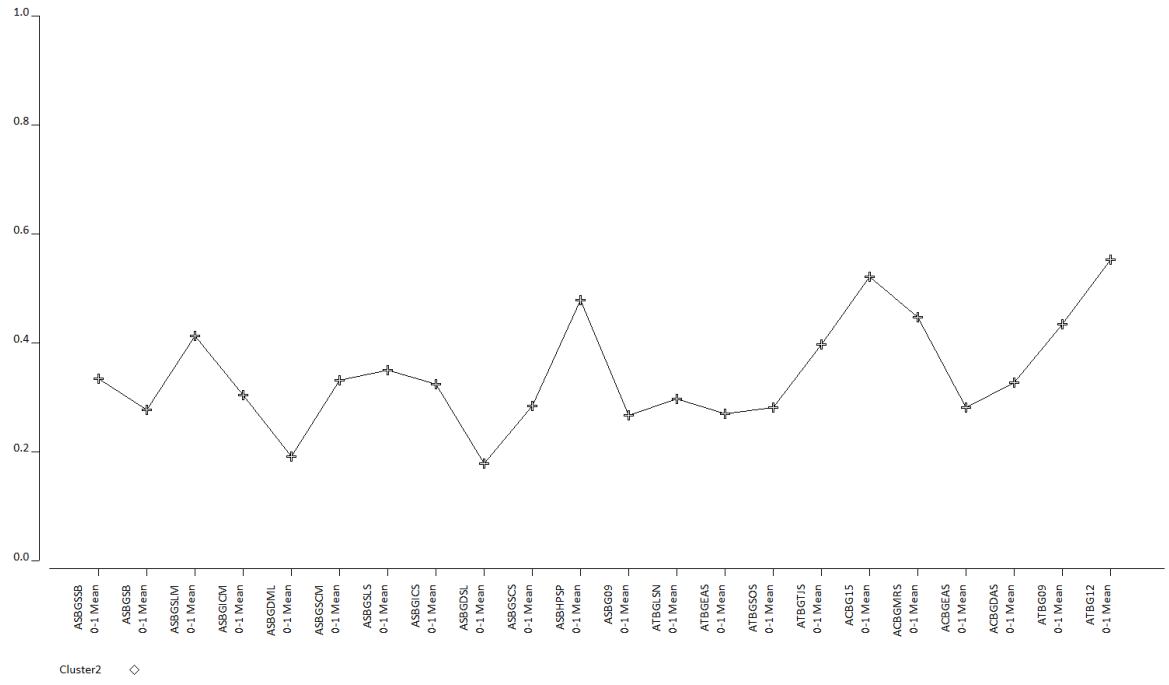
Cluster 2 – Experiencing Difficulties

Schools in Cluster 2 are consistently struggling to maintain school climate health across all respondent types in a way that distinctly separates them from the remaining clusters. For most student variables, Cluster 2 has the lowest overall scores – students on average in these schools do not feel they belong (ASBGSSB), are most routinely bullied (ASBGSB), experience disruption in their math and science lessons (ASBGDML and ASBGDSL, respectively), and do not understand what their teachers are trying to convey (ASBGICS and ASBGICM). While they

are the least confident in mathematics and science (ASBGSCM and ASBGSCS, respectively), they report liking mathematics more than their counterparts in Cluster 1, and their parents are more satisfied than in Cluster 1.

Figure 4.3

MLCA Indicator Means Plot – Cluster 2



Teachers are also reporting more negatively compared to other clusters. Teachers feel their school is not orderly, to the same degree as Cluster 1, they do not report a high emphasis on academics in their school, on average, and report that they the most limited by students not being ready for instruction. Principals report low academic emphasis, problems with attendance amongst teachers, and the lowest sense of discipline.

Variables to examine are students liking mathematics and science and parent satisfaction. Students liking mathematics (ASBGSLM) has a school-level average in Cluster 2 of 9.94; the science equivalent (ASBGSLS) has a value of 9.87. Both correspond to the scale region of "Somewhat like Mathematics" which represents largely positive responses to the items in the

scale. For the parental satisfaction scale, ASBHSPSP has a school-level average of 10.20 for schools in Cluster 2 – a value closer to the means in Cluster 3 and 4 than to the group low in Cluster 1. This is also in the range of "Very Satisfied" – a reminder that overall, the questionnaire data does indicate positive attitudes towards schooling and that these definitions are relative.

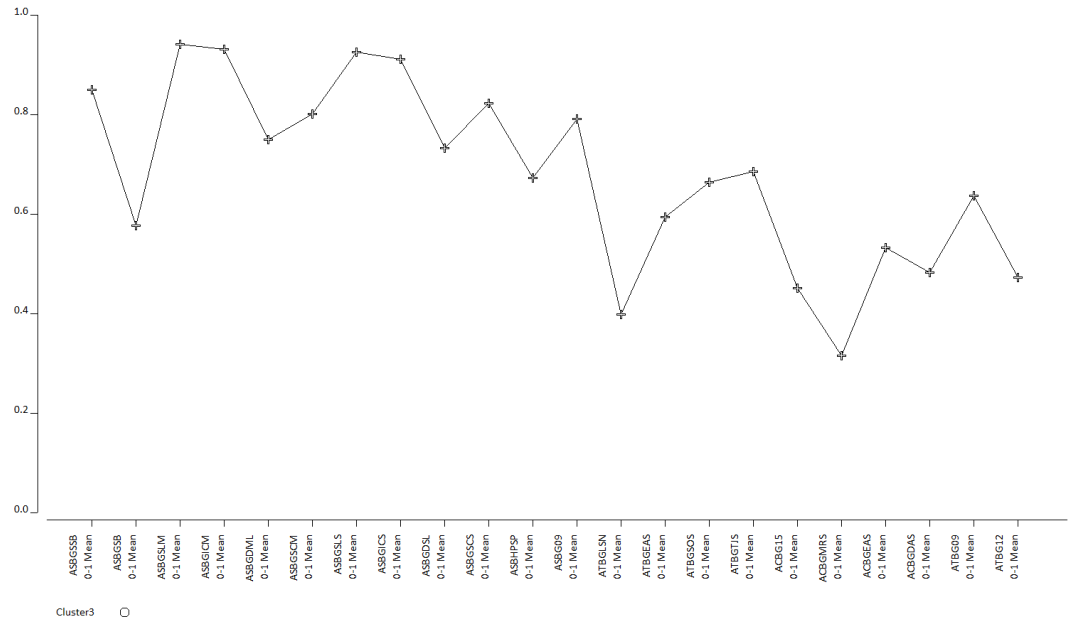
All members of these schools are navigating apparent dysfunction in the operation of a safe and productive school. Students do not feel a strong connection to their school and experience bullying; teachers and principals do not feel that there is a strong emphasis on academics at these schools, although parents indicate on average they are satisfied with the schools.

Cluster 3 – Healthy Social System

Cluster 3 is effectively the "high school climate health category" in that across respondents, schools in this cluster, on average, report that the school system is functioning well. For most of indicator variables, schools in Cluster 3 have the highest mean score.

Figure 4.4

MLCA Indicator Means Plot – Cluster 3



Exceptions are the school average for teachers reporting that students are not ready for instruction (ATBGLSN), the school average of teacher reports of classroom practices (ATBG12), and principal reports of limitations due to research shortages (ACBGMRS). For ATBGLSN, schools in Cluster 3 average 9.39 on the scale which places them in the same scale region as the other clusters – 'some limitations' – but does suggest that teachers report students not being ready, on average, more than in Clusters 1 and 4. ATBG12 is an exception to most of the indicator variables in that high scores are indicative of engaging in the preferred behavior *less*. Thus, teachers in schools in Cluster 3 conduct their classes with engaging practices more often than in schools in Cluster 1 and 2. The principal report suggests that schools in Cluster 3 experience the most limitations due to shortages in mathematics resources.

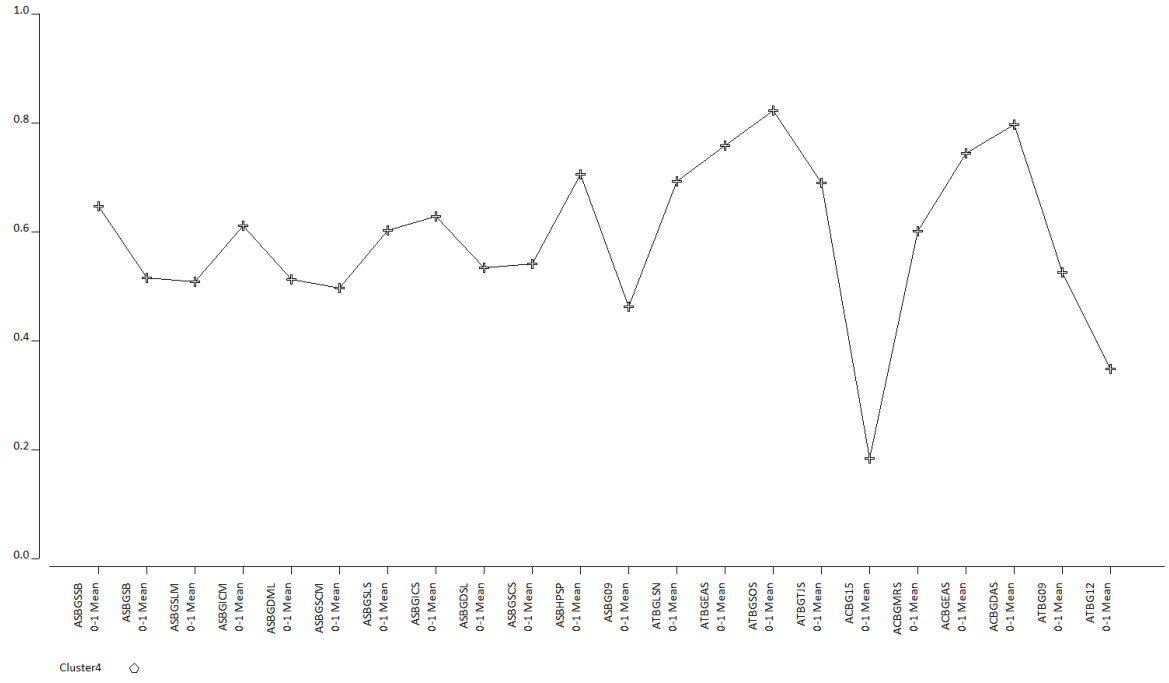
Overall, in the face of some limitations of school readiness and resource shortages, schools in Cluster 3 are positive, productive school environments.

Cluster 4 – Generational Incongruence

Schools in Cluster 4 are judged differently when comparing student points-of-view and those of the adults around them. While the indicators from the student questionnaire for schools in this cluster not low, they are noticeably lower than Cluster 3 whereas the teacher, parent, and principal indicators are the highest of any cluster. In comparison with Cluster 3 where most schools are responding in the "High" response category for most student-level variables, the Cluster 4 response proportions are less positive. For instance, Cluster 3 responses are such that for 9 of 12 student variables, 60% or more schools are in the 'High' response category. Cluster 4 student responses averaged to the school level are less overwhelmingly positive, with the majority of responses in the "Medium" category or evenly split across the three levels. As an illustrative example, 89% of schools in Cluster 3 were in the "High" category for average scores of students liking mathematics (ASBGSLM) compared to just 42% in Cluster 4.

Figure 4.5

MLCA Indicator Means Plot – Cluster 4



While the average student variables were more moderate in Cluster 4, the responses from adults were overwhelmingly positive. Parents in schools in Cluster 4 were the most positive about their schools, teachers were the most satisfied and identified the highest levels of order in the schools. Teachers and principals both identify the greatest emphasis on academics in these schools, and principals also report the highest levels of discipline.

Students in these schools are largely positive about their experience, but the differential in magnitude between their evaluations and those of the adults is a defining characteristic.

Assigning Countries to Level 2 Classes

The MLCA in this study resulted in 3- and 4-class models that were initially suitable based on model fit statistics. However, the 3-class model was ultimately chosen in Step 2b of the analysis to achieve an adequate size of the assigned classes. Hence, the 3-class model was chosen for the subsequent steps of the modeling process, and the final Level-2 solution is summarized in alphabetical order in Table 4.15. Countries were each assigned to exactly one class based on similarities in the proportion of its schools among the four clusters in the final model.

Table 4.15

Country Assignments

	Class 1	Class 2	Class 3
1	Belgium (Flemish)	Albania	Australia
2	Belgium (French)	Armenia	Bosnia & Herzegovina
3	Chinese Taipei	Azerbaijan	Brazil
4	Czech Republic	Bahrain	Canada
5	Denmark	Bulgaria	Chile
6	Finland	Georgia	Cyprus
7	France	Iran	England
8	Germany	Jordan	Ireland
9	Hong Kong SAR	Kosovo	Italy
10	Hungary	Kuwait	Kazakhstan
11	Japan	Montenegro	New Zealand
12	Korea, Rep. of	Morocco	Oman
13	Latvia	North Macedonia	Qatar
14	Lithuania	Portugal	South Africa
15	Macao SAR	Romania	Spain
16	Netherlands	Saudi Arabia	United States
17	Norway	Serbia	
18	Poland	Turkiye	
19	Singapore	United Arab Emirates	
20	Slovak Republic	Uzbekistan	
21	Slovenia		
22	Sweden		

Geographic and cultural similarities are readily apparent within the classes. Class 1 contains most of the Western European countries, Scandinavia, Central Europe, the Baltic countries, and the participating East Asian countries. These countries are part of one Level-2 class based on similarities in the proportions of school climate health clusters within the country. Based on the much written about differences between Western and Eastern conceptualizations of well-being, it is compelling that these countries belong to the same class. At least based on school climate health, the profiles of schools within these countries are more similar internally than compared to countries outside of Class 1.

Turning to Class 2, every participating country from the Middle East is represented, except for Oman and Qatar, as well as most of the participating Balkan countries (except Bosnia & Herzegovina in Cluster 3 and Slovenia in Cluster 1). Neighboring countries Georgia, Azerbaijan, and Armenia are also included within this class. Portugal and Romania are the other two European countries that are not so easily linked to the others in this class. Historical influences such as the imperial impacts of Moors or Byzantine eras or geographic links to other Black Sea nations could underlie the similarities.

Finally, Class 3 has smaller groupings of similar countries compared to the amount of regional similarity covered in Class 1. Most evident is that all the major English-speaking countries are in this class. While other countries do administer in English (e.g. most Middle Eastern participants adapt their instruments to English), this class has all the predominantly English-speaking countries except for Ireland. Furthermore, every Commonwealth country participating in TIMSS is included in this Class (Australia, Canada, England, and New Zealand). There are additional pairs of countries that have surface-level similarities. Chile and Brazil are the only participating countries from South America, and Spain and Italy are both Romance language countries in Western Europe. However, it is more difficult to make any larger groupings out of the countries in this Class or to find any immediate connections between Cyprus or Kazakhstan which would more readily be linked to countries in other Classes.

While external factors such as geography and history can illuminate Class definitions and raises curiosities, this is not the basis for the model definition of these country-level classes. The MLCA conducted in LatentGOLD isolates similarities across countries in the proportion of schools per country in each cluster in the final four-cluster model. Thus, Table 4.16 - 4.18 summarize the percent composition of each country over the three Level-2 classes.

Table 4.16

Country Class 1 School Climate Health Composition

Country	Schools Sampled	% Cluster 1	% Cluster 2	% Cluster 3	% Cluster 4
Belgium (Flemish)	146	53.9	33.4	4.5	8.2
Belgium (French)	165	44.8	47	2.5	5.7
Chinese Taipei	215	83.7	4.7	4.5	7.1
Czech Republic	195	59.4	25.6	5	10
Denmark	158	76.3	14.1	0	9.6
Finland	170	88.3	6.3	1.2	4.2
France	159	63.4	25	4.3	7.4
Germany	230	55.2	34.3	3.9	6.7
Hong Kong SAR	139	75.1	11.6	2.4	10.9
Hungary	154	40.7	33.4	11	15
Japan	141	99.1	0	0	0.9
Korea, Rep. of	156	77	4.2	2	16.8
Latvia	154	49.9	43.7	1.8	4.7
Lithuania	190	41.9	40.2	4.2	13.7
Macao SAR	59	76.3	15.3	1.7	6.8
Netherlands	123	54.8	25.8	4	15.5
Norway	160	70.3	23.8	0.8	5.1
Poland	150	85.5	11.9	0	2.6
Singapore	181	72.9	9.9	0.6	16.6
Slovak Republic	162	44	25.7	15.8	14.5
Slovenia	149	72.7	25.2	1.7	0.4
Sweden	160	78.4	13.8	3	4.8

Country Class 1 Description

Beginning in Class 1, countries in this class share high proportions of their schools in Clusters 1 and 2, with low proportions in both Clusters 3 and 4. This is an intriguing result as Cluster 1 had indications of discord while Cluster 2 was characterized by overall poor school climate health. While the unifying characteristic of the Class is low proportions of Cluster 3 and 4, there is variability in the proportions of Clusters 1 and 2.

The proportion of schools in Cluster 1 is highest in Japan with 99.1% of schools in the sample assigned to Cluster 1, and the lowest proportion is in Hungary with 40.7% of schools. The result in Japan is suggestive of an overwhelmingly uniform school experience at fourth grade, but also warrants caution as these are all aggregated data. For Cluster 2 proportions, Belgium (French) is the highest with 47% followed by Latvia with 43.7%. Countries in this class with the highest proportions of Clusters 1 and 2 leave little room for membership to the other clusters, but the Slovak Republic has the highest percent of schools in the "Healthy Social System" cluster with 15.8% of schools. Considering the characterizations of the Clusters of highest proportion in these countries, the school systems within may be struggling to support a health school climate that engenders positive attitudes and wellbeing among its stakeholders.

Country Class 2 Description

The cluster composition for schools in Class 2 countries is the complement pattern to Class 1. Countries in this class have very few schools in Cluster 1, especially. Many countries in this Class have most of their schools belonging to Cluster 3 – the Healthy Social System Cluster. The country with the highest percentage in this Cluster is Armenia with 90.6% followed by Albania with 86.4%. Schools in Cluster 4 are characterized by largely positive school climate health, especially when considering the points of views of the adults in the schools. Bahrain has the highest proportion of schools in Cluster 4 (45.5%) followed by Portugal (43.4%) and Romania (41.8%).

Table 4.17

Country Class 2 School Climate Health Composition

Country	Schools Sampled	% Cluster 1	% Cluster 2	% Cluster 3	% Cluster 4
Albania	108	0	0	86.4	13.6
Armenia	149	0	0.3	90.6	9.1
Azerbaijan	199	0	18.8	56.1	25.1
Bahrain	187	0	11.8	42.8	45.5
Bulgaria	153	0.5	21	57.8	20.8
Georgia	156	0	4.4	66.8	28.8
Iran	224	0	9.9	72.1	18
Jordan	224	0.5	28.5	56.5	14.5
Kosovo	152	0	6.5	80.8	12.7
Kuwait	150	0	33.9	42.1	24
Montenegro	141	0.7	15.6	55.3	28.4
Morocco	270	0	16.6	79.4	4.1
North Macedonia	149	0	11.5	69.1	19.4
Portugal	164	0.9	22.4	33.3	43.4
Romania	130	0	12.7	45.5	41.8
Saudi Arabia	214	0	19.3	49.7	31.1
Serbia	158	1.4	16.4	51.9	30.3
Turkiye	141	0.2	30.1	45.7	23.9
United Arab Emirates	755	0	22.5	37.7	39.8
Uzbekistan	166	0	2.5	83.1	14.4

Overall, schools in countries in this class are characterized by positive school functioning based on the proportions of schools assigned to Clusters 3 and 4. However, there are still countries with notable percentages of schools in the Experiencing Difficulty Cluster (2). Highest among them is Kuwait with 33.9% in this cluster, followed by Turkiye with 30% and Jordan with 28.5%. An apt reminder that overall judgements can obscure areas where intervention may be needed. However, exploring the variations does not distract from the overall pattern that the unifies countries in Class 2 – most schools in Cluster 3 and Cluster 4.

Country Class 3 Description

Class 3 does not share the overarching two-way split of Class 1 and Class 2. Instead, sharing relatively low representation in Cluster 1, these countries then present variable spread across each of the remaining three school clusters. Most countries, though, the majority of schools are assigned to Cluster 2 and Cluster 4. These two clusters are nearly opposite in characterization; the former defined by low school climate health, and the latter largely positive if variable across respondent. That these are the predominant clusters in this class of countries is an interesting result.

Table 4.18

Country Class 3 School Climate Health Composition

Country	Schools Sampled	% Cluster 1	% Cluster 2	% Cluster 3	% Cluster 4
Australia	283	5.6	52.3	2.7	39.3
Bosnia & Herzegovina	120	6.2	46.5	9.1	38.2
Brazil	796	0.2	56.9	3.9	39
Canada	659	8.3	51.4	9.3	31
Chile	178	1.1	73	4	21.9
Cyprus	148	6.3	48.1	4.3	41.3
England	131	8.8	37.5	0.9	52.7
Ireland	153	18.9	21.4	0.4	59.4
Italy	152	12.3	51.9	10.7	25.1
Kazakhstan	185	5.5	6.1	10.7	77.6
New Zealand	154	2.6	72	4.3	21.1
Oman	220	0.2	49.6	16	34.3
Qatar	279	1.4	20.8	16.5	61.3
South Africa	285	2.9	55.1	11.9	30
Spain	524	2.9	37.1	4.3	55.7
United States	241	4.3	53.5	3.2	39

Within the class, countries vary in where most schools are assigned. Countries that have a majority in Cluster 4 include England (52.7%), Ireland (59.4%), Kazakhstan (77.6%), Qatar (61.3%), and Spain (55.7%) still have large proportions of their schools experiencing a stark

contrast in school experience in Cluster 2. More countries in this Class have the majority of their schools assigned to Cluster 2; Brazil is highest with 56.9% followed by South Africa (55.1%) and the United States (53.5). While the other country-level classes also evidenced two-way splits between Clusters, Class 3 suggests a clear divide in quality of education between its majority parts.

Towards introducing nuance into these groups, the temptation to label groups of countries based on their most present school climate health type is discouraged. Looking at the country classes overall, there are schools of each school climate health cluster in countries with a large representation of other clusters (see Table 4.19). Consider that 15.2% of schools are struggling to sustain school climate health (Cluster 2) even though they are in countries with the most schools excelling in the same arena (Class 2). Class 3 may have the majority of its schools in Cluster 2, but there are nevertheless roughly 5% of schools in these countries sustaining healthy school climates.

Table 4.19

Country Classes by School Cluster Proportions

Country Class	School Climate Health Cluster	Percent of Total Schools in Class
1	1	69.9
	2	19.9
	3	3.2
	4	7.0
2	1	0.1
	2	15.2
	3	65.4
	4	19.3
3	1	3.6
	2	51.4
	3	4.7
	4	40.3

Table 4.20 summarizes similar information, but based on the distribution of the total schools by school climate health cluster over the three country Classes as an alternate approach to these results. Consider the cluster with the lowest overall school climate health, Cluster 2. While most of these schools reside in Class 3, 10-15% of the schools assigned to this Cluster are in countries in the other two Classes. These results evidence that some schools are still struggling in their social environment that deserve the spotlight for potential innovation and to not be shrouded by the overall success in that country Class.

Table 4.20

School Clusters by Proportion Assigned per Country Class

School Climate Health Cluster	Country Class	Percent of Total Schools in Cluster
1	1	91.0
	2	.1
	3	8.9
2	1	15.1
	2	10.9
	3	74.1
3	1	4.3
	2	83.6
	3	12.1
4	1	6.9
	2	17.8
	3	75.3

Secondary Analyses

Previous MLCA studies have conducted secondary analyses to investigate the nature of relationships between group membership and possibly influential variables. Measures of socioeconomic status are commonly included, as in clusters extracted from the PISA Global Competence assessment (Lee & Stankov, 2023) and student-level school climate (Allison et al., 2016).

The purpose of the secondary analyses is to conduct a preliminary investigation into the possible influence of confounding variables in characterizing school clusters and to understand the relationship between cluster membership and mathematics achievement. Following directly from the theoretical scrutiny in capturing concepts relating to school climate and wellbeing, students with cultural backgrounds different from the predominant one ought to be considered.

Description of Variables and Analysis Setup

Six variables from the TIMSS 2023 Context Questionnaires were chosen for these secondary analyses as they are responsive to the literature and available from the school questionnaire such that no aggregation was necessary. For reference, an overview of the six school-level variables from the TIMSS 2023 School Questionnaire is included in Table 4.21

Table 4.21

Overview of Variables Selected for Covariate Analysis

Variable	Item name	Categories
ACBG02	Total G4 student enrollment	n/a
ACBG03A	Percentage of students from disadvantaged background	4
ACBG04	Percentage of students speak the language of test at home	5
ACBG05A	How many people are in the surroundings	7
ACBG05B	Urban rural etc.	5
ACBG15	School readiness	Max: 44

Differences across school climate health clusters are apparent in the descriptive statistics presented in Table 4.22. Clusters 2 and 3 have the largest school sizes of Grade 4 students (53.13 and 52.35, respectively), although there is greater variation in size denoted by the larger standard deviation. Cluster 2 has the largest proportion of students from disadvantaged backgrounds (3.1) indicating the mean school in cluster 2 has between 26% and 50% of students disadvantaged based on principal reports. Cluster 1 has the lowest mean score (1.85) on that variable (ACBG03A) with the mean school proportion of students from disadvantaged students of less

than 10%. The clusters are relatively even for ACBG04 suggesting linguistically homogenous student populations, although Clusters 2 and 3 have mean scores suggesting greater proportions of multilingual students than the other two clusters. Schools in Clusters 1 and 4 have the most prepared students per principal reports (33.75 and 35.18, respectively) out of the maximum score of 44.

Table 4.22

Descriptive Statistics – School Characteristics by Cluster in Secondary Analyses

	1		2		3		4	
	Mean	(s.d)	Mean	(s.d)	Mean	(s.d)	Mean	(s.d)
ACBG02	46.28	46.33	53.13	60.69	52.34	74.43	42.15	61.20
ACBG03A	1.85	0.95	3.10	1.09	2.59	1.22	2.25	1.12
ACBG04	1.44	0.91	1.94	1.44	1.99	1.60	1.67	1.22
ACBG05A	4.46	2.26	4.38	2.19	4.74	2.29	4.51	2.22
ACBG05B	3.33	1.26	2.98	1.42	3.24	1.40	3.04	1.41
ACBG16	33.75	10.29	28.61	11.71	27.52	12.46	35.18	11.47

Recodes were implemented to support the interpretation of the analysis results. First, the aim was to create relatively equal subgroups, which was mostly successful. An exception was ACBG04, where the vast majority of schools reported more than 90% of students spoke the language of the test at home. An additional consideration was to describe a sensible intercept school i.e. that the characteristics of the comparison case for the variables would combine to describe a realistic school. The recodes were executed as in Table 4.22. Descriptive statistics following the recoding of the analysis variables are included in Table 4.23

Table 4.23

Recode Scheme for Secondary Analysis Variables

Variable	Recode	Category labels
ACBG02	0-50 → 2	2 – Relatively small
	51-100 → 1	1- Somewhat Large
	101+→ 0	0- Large
ACBG03	1 (0-10%)→ 2	2 - 0-10% Disadvantaged
	2 (11-25%) → 1	1- 11-50% Disadvantaged
	3 (26-50%) → 1	0- More than 50%
	4 (more than 50%) → 0	Disadvantaged
ACBG04	1 (more than 90%) → 1	1 - 90% Speak Language of
	2 (76-90%) → 0	Test
	3 (51-75%) → 0	0 - Less than 90% Speak
	4 (26-50%) → 0	Language of Test
	5 (25% or less)→ 0	
ACBG05A	1 (More than 500,000 people) → 0	0 - Large Population
	2 (100,001 to 500,000 people) → 0	1 - Medium Population
	3 (50,001 to 100,000 people) → 0	2-Small Population
	4 (30,001 to 50,000 people) → 1	
	5 (15,001 to 30,000 people) → 1	
	6 (3,001 to 15,000 people) → 1	
	7 (3,000 people or fewer) → 2	
ACBG05B	1 Urban–Densely populated → 0	0- Urban/Suburban
	2 Suburban–On fringe or	1- Rural/Small town
	outskirts of urban area → 0	
	3 Medium size city or large town → 0	
	4 Small town or village → 1	
	5 Remote rural→ 1	
ACBG15	Based on the summed score	
	0-22 → 2	2- Students Less Prepared
	23-38→ 1	1- Students Mostly Prepared
	38+→ 0	0- Students More Prepared

Table 4.24

Descriptive Statistics Following Recode

	1		2		3		4	
	Mean	(s.d)	Mean	(s.d)	Mean	(s.d)	Mean	(s.d)
ACBG02	1.58	0.64	1.51	0.70	1.56	0.70	1.65	0.63
ACBG03A	1.38	0.63	0.63	0.71	0.94	0.79	1.13	0.72
ACBG04	0.75	0.43	0.64	0.48	0.69	0.46	0.71	0.45
ACBG05A	0.89	0.83	0.85	0.79	1.03	0.82	0.91	0.79
ACBG05B	0.56	0.50	0.44	0.50	0.51	0.50	0.47	0.50
ACBG16	0.77	0.75	1.11	0.79	1.18	0.81	0.71	0.77

Investigation of Possible Confounding Variables

Interpretation of this analysis hinges on the assumption that a change in school characteristics would not influence the probability of cluster assignment. Findings of such influence suggest confounding similarities amongst schools within the same cluster other than school climate health alone. This is an important facet to understand as a point of validity that the groupings from the MLCA represent more than just demographic similarity

For this analysis, the basis of comparison is a school with less than 50 Grade 4 students located in rural surroundings of less than 3,000 inhabitants. Students in such a school are the least prepared at the start of school, but the majority (90%) of the student body is not disadvantaged and speaks the language of the test at home. With this case in mind, comparisons of the likelihood of cluster memberships can be made based on individual changes from this base case. Cluster 2 (Experiencing Difficulties) was chosen as the reference cluster for this analysis such that characteristics relating to the likelihood of belonging to the lower school climate health cluster can be better understood. Results are organized by cluster; odds are interpreted from the Exp(B) column in each table.

The odds of belonging to Cluster 1 (System Misalignment) compared to Cluster 2 are summarized in Table 4.25. Notably, an increase in the readiness of students (ACBG15) from

either degree increases the odds of being in Cluster 1 rather than Cluster 2 by the largest amount (68% increase from 2 to 0; 81% increase from 2 to 1). The most substantial decrease in odds is for percent disadvantaged (ACBG03A) - whereas for schools with more than 50% of its students disadvantaged the odds of belonging to Cluster 1 decrease by 86% compared to schools with less than 90% disadvantaged. Also notable is the decrease in odds by 56% of being in Cluster 1 compared to Cluster 2 for schools in urban or suburban settings (ACBG05B) compared to remote or rural settings. The odds of being in Cluster 1 compared to Cluster 2 also decrease for ACBG04; schools with less than 90% of students speaking the language of the test at home (33%) and for schools with more than 100 grade 4 students (49%).

Table 4.25

Cluster Comparisons–Cluster 1 to Cluster 2

	B	Std. Error	Wald	df	Sig.	Exp(B)	Lower Bound	Upper Bound
Intercept	-10.05	0.05	35068.35	1.00	0.00			
[ACBG02=0]	-0.67	0.02	1193.04	1.00	0.00	0.51	0.49	0.53
[ACBG02=1]	-0.15	0.01	124.56	1.00	0.00	0.86	0.84	0.88
[ACBG03A=0]	-1.96	0.02	11975.69	1.00	0.00	0.14	0.14	0.15
[ACBG03A=1]	-0.60	0.01	2172.67	1.00	0.00	0.55	0.54	0.56
[ACBG04=0]	-0.40	0.01	1090.53	1.00	0.00	0.67	0.66	0.69
[ACBG05A=0]	0.03	0.02	2.38	1.00	0.12	1.03	0.99	1.07
[ACBG05A=1]	-0.36	0.01	567.32	1.00	0.00	0.70	0.68	0.72
[ACBG05B=0]	-0.83	0.01	3127.24	1.00	0.00	0.44	0.42	0.45
[ACBG16=0]	0.52	0.01	1355.45	1.00	0.00	1.68	1.64	1.73
[ACBG16=1]	0.59	0.01	1851.00	1.00	0.00	1.81	1.76	1.86

The odds of belonging to Cluster 3 (Healthy Social System) compared to Cluster 2 are summarized in Table 4.26. Comparing these two clusters, there were no factors that would increase the odds of a school being assigned to Cluster 3 than to Cluster 2. The two factors with the least magnitude of decrease were schools with more than 100 Grade 4 students (ACBG02; 7% decrease) and urban settings (ACBG05B; 11% decrease). The greatest magnitude decrease in

the odds of being in Cluster 3 compared to Cluster 2 was having more than 50% of students disadvantaged (ACBG03A; 47% decrease) and being in a city or town with more than 50,000 residents (ACBG05A; 42% decrease). There was also a reduction of about 30-40% in the odds of being in Cluster 3 for an increase in the proportion of the school ready for instruction (ACBG15). Overall, the magnitude of the comparative difference was more subdued than in the results comparing the likelihood between Cluster 1 and Cluster 2.

Table 4.26

Cluster Comparisons–Cluster 3 to Cluster 2

	B	Std. Error	Wald	df	Sig.	Exp(B)	Lower Bound	Upper Bound
Intercept	-4.43	0.04	10710.02	1.00	0.00			
[ACBG02=0]	-0.07	0.02	20.13	1.00	0.00	0.93	0.90	0.96
[ACBG02=1]	-0.32	0.01	648.10	1.00	0.00	0.73	0.71	0.74
[ACBG03A=0]	-0.63	0.01	2004.63	1.00	0.00	0.53	0.52	0.55
[ACBG03A=1]	-0.49	0.01	1333.68	1.00	0.00	0.61	0.60	0.63
[ACBG04=0]	-0.36	0.01	1125.39	1.00	0.00	0.70	0.68	0.71
[ACBG05A=0]	-0.54	0.02	1089.48	1.00	0.00	0.58	0.56	0.60
[ACBG05A=1]	-0.44	0.01	1058.97	1.00	0.00	0.64	0.63	0.66
[ACBG05B=0]	-0.11	0.01	73.15	1.00	0.00	0.89	0.87	0.92
[ACBG16=0]	-0.51	0.01	1705.61	1.00	0.00	0.60	0.58	0.61
[ACBG16=1]	-0.42	0.01	1364.06	1.00	0.00	0.66	0.64	0.67

The odds of belonging to Cluster 4 (Generational Incongruence) compared to Cluster 2 are summarized in Table 4.27. The comparison between these two clusters resulted in the greatest magnitude increase in odds of the three cluster comparisons. For schools with students very ready compared to those with students not ready (ACBG15), the odds of being in Cluster 4 compared to Cluster 2 increase by 147%; for students mostly ready the increase in odds is 49%. There is also a 10% increase in the odds of being in Cluster 4 compared to Cluster 2 for schools in urban or suburban environments compared to small towns (ACBG05B). The factor with the greatest odds decrease was the proportion of students disadvantaged (ACBG03A) for which a

school with more than 50% of students disadvantaged has a 73% decrease in odds of being in Cluster 4 rather than Cluster 2. There was also a 46% decrease in odds for schools larger than 100 Grade 4 students compared to smaller schools (ACBG02). The comparison between Cluster 4 and Cluster 2 had the smallest related decrease in odds for an increase in the proportion of students speaking the language of the test at home (ACBG04). For schools with less than 90% of students speaking the language of the test at home, there was only a 20% decrease in odds compared to a 30% decrease in the other two comparisons.

Table 4.27

Cluster Comparisons–Cluster 4 to Cluster 2

	B	Std. Error	Wald	df	Sig.	Exp(B)	Lower Bound	Upper Bound
Intercept	-3.20	0.04	6915.37	1.00	0.00			
[ACBG02=0]	-0.62	0.02	1492.00	1.00	0.00	0.54	0.52	0.56
[ACBG02=1]	-0.47	0.01	1627.56	1.00	0.00	0.63	0.61	0.64
[ACBG03A=0]	-1.30	0.01	9422.96	1.00	0.00	0.27	0.27	0.28
[ACBG03A=1]	-0.45	0.01	1445.85	1.00	0.00	0.64	0.62	0.65
[ACBG04=0]	-0.23	0.01	523.17	1.00	0.00	0.79	0.78	0.81
[ACBG05A=0]	-0.36	0.02	559.93	1.00	0.00	0.70	0.68	0.72
[ACBG05A=1]	-0.18	0.01	212.92	1.00	0.00	0.83	0.81	0.85
[ACBG05B=0]	0.10	0.01	68.54	1.00	0.00	1.10	1.08	1.13
[ACBG16=0]	0.90	0.01	6026.52	1.00	0.00	2.47	2.41	2.52
[ACBG16=1]	0.40	0.01	1204.81	1.00	0.00	1.49	1.46	1.53

Overall, there is indication that school characteristics outside of the school climate health definition are perhaps confounding the classification of these clusters. In other words, other factors might explain the similarities within the four defined clusters.

Achievement Analysis

Examining relationships between school cluster membership and average school achievement is not the primary justification for the present research but is nonetheless worthy of investigation. Estimating trends in average mathematics and science achievement is among the

major goals of the TIMSS assessment and is valuable data for systems-level evaluation of education quality internationally. Understanding the relationship between school climate health clusters and mathematics achievement increases the utility of the results and elaborates the interplay between the variables.

Mathematics achievement as measured by the TIMSS assessments influences policy internationally and provides critical information on the extent that national education systems are successfully educating their children. As such a trusted measure, along with trends reaching as far back as 1995, finding that schools with high school climate health support increased achievement scores could bolster interventions and increase the profile of the qualities therein. However, there could be other mechanisms at work in the relationship between school climate health and achievement, so the variables from the previous analysis are also included as described in the previous section.

Comparisons Across Clusters. Cluster 2 (Experiencing Difficulty) is once again the reference group for this analysis. The analysis was conducted with cluster assignment as a factor alongside those variables in an analysis of variance with school mean Grade 4 mathematics achievement (referred to moving forward as, Achievement) as the outcome. The parameter estimates of the main effects resulting from the analysis are summarized in Table 4.28.

Table 4.28

Summary Results Table – Achievement Analysis Main Effects

	B	Std. Error	t	Sig.	Lower Bound	Upper Bound
Intercept	457.11	0.56	810.10	0.00	456.01	458.22
Cluster 1	82.24	0.85	96.83	0.00	80.57	83.90
Cluster 3	49.37	0.83	59.63	0.00	47.75	50.99
Cluster 4	40.51	0.80	50.76	0.00	38.95	42.08
[ACBG02=0]	13.39	0.52	25.73	0.00	12.37	14.41
[ACBG02=1]	12.46	0.39	31.95	0.00	11.69	13.22
[ACBG03A=0]	-54.59	0.49	-110.62	0.00	-55.56	-53.62
[ACBG03A=1]	-14.57	0.51	-28.77	0.00	-15.56	-13.58
[ACBG04=0]	41.02	0.34	121.89	0.00	40.36	41.68
[ACBG05A=0]	6.35	0.52	12.12	0.00	5.32	7.38
[ACBG05A=1]	11.96	0.46	26.23	0.00	11.07	12.85
[ACBG05B=0]	-5.50	0.40	-13.59	0.00	-6.29	-4.71
[ACBG16=0]	18.15	0.41	44.27	0.00	17.35	18.95
[ACBG16=1]	5.28	0.37	14.39	0.00	4.56	6.00

The intercept conditions for this analysis describe a school in Cluster 2 with less than 50 Grade 4 students located in rural surroundings of less than 3,000 inhabitants. Students in this kind of school are the least prepared at the start of school, but the majority (90%) of the student body is not disadvantaged and speaks the language of the test at home. The predicted achievement for a school with these characteristics is 457.11 (found in 'B' column of each table), which is within the Low Benchmark for mathematics, the lowest of four benchmarks categorizing international achievement (see Mullis et al., 2020). Membership in any of the other clusters results in an increase in predicted Achievement, with an increase of 40.51 for Cluster 4; 49.37 for Cluster 3; and 82.24 for Cluster 1. While it is a compelling result to see an 82-point difference between Cluster 2 and Cluster 1, it would not be appropriate to claim this was directly the result of school climate health conditions given the evidenced skew in distribution of clusters across countries.

Continuing with the main effects, there is also a predicted increase in Achievement for schools of increased size in Cluster 2 (ACBG02; 13.39 for schools with more than 100 students; 12.46 for schools with between 51 and 100), as well as for schools with more populous environments (ACBG05A; 11.96 for midsize populations, 6.35 for cities with over 50,000), and for increases in the extent that students are prepared for school (ACBG15; 5.28 increase for mostly prepared, 18.15 for very prepared). There is a notable decrease in predicted achievement for schools with more disadvantaged students (ACBG03A) than the comparison with a 54.59 decrease for schools where more than 50% of students are disadvantaged and 14.57 for schools with between 11 and 50% disadvantaged.

The interaction parameter estimates for Cluster 1 in comparison to the Cluster 2 reference group are summarized in Table 4.29. A note to the interpretation that the coefficients for all interaction terms are in addition to the increase in achievement score predicted by a change in cluster membership alone.

Table 4.29

Cluster 1 Comparison – Achievement Analysis

	B	Std. Error	t	Sig.	Lower Bound	Upper Bound
[CLUSTER=1.00] * [ACBG02=0]	4.80	0.95	5.04	0.00	2.93	6.66
[CLUSTER=1.00] * [ACBG02=1]	-7.02	0.67	-10.52	0.00	-8.33	-5.72
[CLUSTER=1.00] * [ACBG03A=0]	0.72	0.96	0.76	0.45	-1.16	2.60
[CLUSTER=1.00] * [ACBG03A=1]	-2.01	0.67	-2.99	0.00	-3.32	-0.69
[CLUSTER=1.00] * [ACBG04=0]	-55.20	0.62	-89.21	0.00	-56.41	-53.98
[CLUSTER=1.00] * [ACBG05A=0]	31.51	0.92	34.40	0.00	29.72	33.31
[CLUSTER=1.00] * [ACBG05A=1]	1.04	0.74	1.41	0.16	-0.40	2.49
[CLUSTER=1.00] * [ACBG05B=0]	9.02	0.77	11.77	0.00	7.51	10.52
[CLUSTER=1.00] * [ACBG16=0]	-18.97	0.73	-26.04	0.00	-20.40	-17.54
[CLUSTER=1.00] * [ACBG16=1]	-15.66	0.70	-22.46	0.00	-17.02	-14.29

With that in mind, we see that two effects on Achievement in this model are not statistically significant at an alpha error level of 0.05 – an increase in the proportion of students that are disadvantaged (ACBG03A; p=0.45) and schools in midsize towns and cities

(ACBG05A; p=0.16). Predicted decreases in Achievement as school preparedness increases is an odd result that is possibly an artifact of high preparedness overall such that the low preparedness case is not well-estimated. The interaction between urbanicity (ACBG05A) and cluster membership is also interesting with a predicted increase in Achievement of 31.51 in schools in Cluster 1 in large cities with more than 50,000 residents. The predicted influence of the proportion of students in school not speaking the language of the test at home (ACBG04) is compelling in contrast to Cluster 2. In that cluster, there was about a 40-point increase whereas in Cluster 1 there is a predicted 55.20 decrease in predicted Achievement which reduces the predicted advantage of being in Cluster 1 compared to Cluster 2 to just 27 points.

The interaction parameter estimates for Cluster 3 in comparison to the Cluster 2 reference group are summarized in Table 4.30. There is a lack of alignment in interpreting some of the parameter estimates, pointing to possible limitations in the design such as the coding scheme. For instance, there is a predicted Achievement increase of 26.89 for school surroundings that are not rural (ACBG05B), but a reduction in predicted Achievement for both the comparison to midsize cities and towns (35.93) and large cities (17.51), both ACBG05A. Similar to Cluster 1, there is a decrease in predicted Achievement for schools with greater proportions of students not speaking the language of test at home (ACBG04) with a magnitude (38.57) great enough to mostly eliminate the advantage of being in Cluster 3. This is a particularly meaningful result as Cluster 3 and Cluster 2 represent the extreme cases – and multilingual students do not seem to reap the benefits of a healthier school climate.

Table 4.30

Cluster 3 Comparison – Achievement Analysis

	B	Std. Error	t	Sig.	Lower Bound	Upper Bound
[CLUSTER=3.00] * [ACBG02=0]	-9.31	0.86	-10.80	0.00	-10.99	-7.62
[CLUSTER=3.00] * [ACBG02=1]	-7.83	0.68	-11.49	0.00	-9.16	-6.49
[CLUSTER=3.00] * [ACBG03A=0]	36.23	0.76	47.89	0.00	34.74	37.71
[CLUSTER=3.00] * [ACBG03A=1]	9.11	0.74	12.30	0.00	7.66	10.56
[CLUSTER=3.00] * [ACBG04=0]	-38.57	0.58	-66.47	0.00	-39.71	-37.44
[CLUSTER=3.00] * [ACBG05A=0]	-17.51	0.88	-19.92	0.00	-19.23	-15.78
[CLUSTER=3.00] * [ACBG05A=1]	-35.93	0.74	-48.48	0.00	-37.38	-34.48
[CLUSTER=3.00] * [ACBG05B=0]	26.89	0.71	38.05	0.00	25.50	28.27
[CLUSTER=3.00] * [ACBG16=0]	-9.30	0.68	-13.67	0.00	-10.64	-7.97
[CLUSTER=3.00] * [ACBG16=1]	-10.58	0.62	-17.06	0.00	-11.80	-9.37

The interaction parameter estimates for Cluster 4 in comparison to the Cluster 2 reference group are summarized in Table 4.31. Overall, the coefficients are not of as great magnitude as in the comparisons with Cluster 1 and 2. The most impactful are school size and school locality. In the case of the former, there is a predicted increase of 28.14 for schools larger than 100 students and 19.87 for schools between 50 and 100 (ACBG02). Considering school locality, there is a 19.13-point decrease in Achievement for schools in towns and cities between 3000 and 50000 people (ACBG05A). In contrast to what was observed in Cluster 3, there is a decrease predicted both as the proportion of disadvantaged students (ACBG03A) increases (8.09) and as the proportion of students not speaking the language of the test (ACBG04) increases (4.97).

Table 4.31

Cluster 4 Comparison – Achievement Analysis

	B	Std. Error	t	Sig.	Lower Bound	Upper Bound
[CLUSTER=4.00] * [ACBG02=0]	28.14	0.86	32.61	0.00	26.45	29.83
[CLUSTER=4.00] * [ACBG02=1]	19.87	0.62	31.84	0.00	18.65	21.09
[CLUSTER=4.00] * [ACBG03A=0]	-2.30	0.72	-3.21	0.00	-3.70	-0.89
[CLUSTER=4.00] * [ACBG03A=1]	-8.09	0.65	-12.37	0.00	-9.37	-6.81
[CLUSTER=4.00] * [ACBG04=0]	-4.97	0.53	-9.37	0.00	-6.01	-3.93
[CLUSTER=4.00] * [ACBG05A=0]	-2.17	0.81	-2.67	0.01	-3.75	-0.58
[CLUSTER=4.00] * [ACBG05A=1]	-19.13	0.69	-27.77	0.00	-20.48	-17.78
[CLUSTER=4.00] * [ACBG05B=0]	2.44	0.63	3.91	0.00	1.22	3.67
[CLUSTER=4.00] * [ACBG16=0]	-10.65	0.65	-16.49	0.00	-11.91	-9.38
[CLUSTER=4.00] * [ACBG16=1]	-13.21	0.64	-20.78	0.00	-14.46	-11.97

Comparisons across Country, Class, and Cluster

The results from the analysis of variance suggest that variation in achievement warrants investigation into the interaction of cluster and country where the majority of each cluster resides rather than of the school climate health characteristics alone. This analysis aims at interested in understanding the influence of better school climate health conditions on learning as it exists in schools internationally. Given that achievement is not constant across countries, and cluster composition varies across countries as well, it is important to investigate the interaction of these groupings. Table 4.32 summarizes actual school mean achievement over cluster, country class, and their intersection.

Table 4.32

Multiple Comparisons of Achievement

Cluster Means

	Mean	(s.d.)	Schools (weighted)
1	540.31	46.82	89790
2	455.32	67.94	154565
3	500.32	62.83	86251
4	495.94	71.76	119252

Country Class Means

	Mean	(s.d)	Schools (weighted)
1	528.66	52.71	116815
2	493.03	66.23	110286
3	471.63	74.38	222758

Cluster by Country Class

Country Class	1				2				3			
Cluster	1	2	3	4	1	2	3	4	1	2	3	4
Achievement	540.51	482.02	526.43	544.08	540.89	465.40	499.15	493.86	538.22	448.41	499.22	492.06
s.d	46.19	50.72	51.06	45.19	53.43	67.07	62.26	72.75	52.72	69.58	68.24	71.92
Schools (weighted)	81673	23292	3681	8169	98	16804	72130	21254	8019	114469	10440	89829

Cluster 2 has the lowest school mean achievement with 455.32, while Cluster 1 has the highest at 540.31; Class 1 has the highest mean at 528.66, and Class 3 has the lowest at 471.63. However, recalling the results of the MLCA, a large proportion of Cluster 1 resides in the highest-performing countries such that the high-performing schools are possibly succeeding not because of the qualities of the climate but because of the education system they most commonly occur in. Class 1 has the highest mean school mathematics achievement at 528.66, but within that, followed by Class 2 with 493.03, and Class 3 with 471.63. More information is gleaned from the interaction of Class and cluster. While Class 1 has the highest score overall, it is still true that Cluster 2 schools have the lowest mean achievement within the class (482.02). In fact, Cluster 2 has the lowest mean achievement in all three country classes. However, it is also clear from the results as presented in Table 4.32 that Cluster 3 is the second highest in Class 2 and Class 3 (499.15 and 499.22, respectively). There is evidence both for the deleterious effect of being in a school with overall poor functioning and the possible benefit of sustaining school

climate health to the academic attainment. The intersection between Country Class and School Climate Health Cluster helps understand the nature of the relationship between school mean mathematics achievement and the results of this dissertation.

At the country level, the evidence is more striking. Table 4.33 summarizes the actual mean Grade 4 mathematics achievement within school climate health cluster for each country included in this analysis, organized by the three country classes.

Table 4.33

Country Mean Mathematics Achievement by School Climate Health Cluster

Country Class 1

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Belgium (Flemish)	532.71	490.90	528.24	517.85
Belgium (French)	504.43	473.39	484.74	508.13
Chinese Taipei	594.99	569.33	598.34	603.41
Czech Republic	531.89	507.75	591.55	559.87
Denmark	527.47	489.26	--	518.03
Finland	526.56	494.27	520.27	535.31
France	493.90	458.06	506.56	504.35
Germany	532.34	484.42	542.78	559.38
Hong Kong SAR	582.24	524.19	623.26	623.05
Hungary	515.60	469.08	543.86	542.75
Japan	586.08	--	--	612.48
Korea, Rep. of	581.85	536.44	612.96	602.16
Latvia	530.94	505.06	526.44	525.57
Lithuania	557.89	519.93	529.30	552.84
Macao SAR	570.28	517.72	639.99	589.74
Netherlands	544.03	512.84	528.82	545.47
Norway	535.66	517.23	539.57	556.75
Poland	542.26	517.54	--	538.91
Singapore	612.09	574.84	626.06	621.62
Slovak Republic	530.77	480.27	482.83	531.83
Slovenia	515.44	496.35	525.71	532.42
Sweden	537.41	490.97	476.26	539.78

Country Class 2

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Albania	--	--	523.28	510.71
Armenia	--	445.93	527.39	491.80
Azerbaijan	--	422.72	497.02	465.51
Bahrain	--	429.11	484.94	457.22
Bulgaria	536.62	426.59	483.85	504.61
Georgia	--	471.93	512.87	492.86
Iran	--	456.91	502.57	491.67
Jordan	453.27	402.40	450.43	445.13
Kosovo	--	365.77	444.85	424.71
Kuwait	--	480.50	504.54	505.19
Montenegro	585.30	452.98	494.69	479.23
Morocco	--	461.81	517.45	500.09
North Macedonia	--	453.90	484.03	469.50
Portugal	560.96	489.20	530.45	536.05
Romania	--	506.76	542.57	530.02
Saudi Arabia	--	388.86	449.98	402.11
Serbia	541.08	492.04	509.98	523.14
Turkiye	589.18	531.32	530.88	585.62
United Arab Emirates	--	427.99	526.45	503.32
Uzbekistan	--	394.72	446.57	418.85

Country Class 3

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Australia	555.96	496.27	560.40	542.27
Bosnia & Herzegovina	484.30	426.92	496.94	450.27
Brazil	392.17	382.17	439.17	421.32
Canada	518.32	483.78	521.16	511.25
Chile	424.85	427.46	449.30	462.63
Cyprus	505.56	489.00	521.92	504.20
England	586.13	528.80	557.90	562.49
Ireland	574.16	516.42	527.85	550.86
Italy	554.70	500.22	520.61	528.84
Kazakhstan	472.90	483.21	503.60	487.06
New Zealand	540.86	473.82	527.15	521.57
Oman	393.40	412.61	485.99	451.18
Qatar	532.28	443.69	478.63	474.33
South Africa	447.69	479.62	493.66	519.17
Spain	493.66	480.31	478.68	504.98
United States	543.58	488.39	573.08	529.38

There is no country where Cluster 2 has the highest mean achievement, whereas nearly half of countries (28 of 58) have their highest Achievement in schools assigned to Cluster 3. In 50 of these countries, the school cluster with the lowest Achievement is Cluster 2; the average difference between high and low is 57.06 points across the countries included in this analysis. Countries where Cluster 3 has the highest school mean mathematics achievement are shaded in grey. Notably three countries with the highest achievement in Grade 4 mathematics in TIMSS 2019 (Hong Kong, Korea, and Singapore) are among the countries for whom the schools with Healthy Social Systems are the highest scoring (623.26, 612.96, and 626.06, respectively).

These results are compelling as a whole and suggest an overall pattern that across countries, schools in healthier school climate clusters are observed to have higher mean Grade 4 TIMSS mathematics achievement than those in poor school climate health clusters. However, it is necessary to reflect back on the previously discussed proportions of schools in each cluster by country and take care in making inferences at the country level. For instance, in country Class 1 there are seven countries with their highest mean school achievement in Cluster 3 (Czech Republic, France, Hong Kong, Macao, and Singapore). Referring to Table 4.16 these countries have, respectively, 5, 4.3, 2.4, 1.7, and 0.6 percent of their schools that participated in TIMSS in Cluster 3 (*Healthy Social System*); this represents less than 10 schools per country. Furthermore, considering the potential role of demographic characteristics on the likelihood of school climate health cluster assignment, these few schools could be otherwise advantaged towards high achievement. In summation, an intriguing overall result requires further investigation at the local level to further expand understanding of possible relationships between school environmental conditions and achievement.

CHAPTER 5 - DISCUSSION

This dissertation was designed to respond to the evident need for further definition around desirable, attainable states of school social environments internationally. Children are entitled to quality education and the characteristics of quality that education systems must therefore sustain have expanded beyond measures of academic achievement alone. This research responds to the call to action from UNESCO in the recent publication of the global Happy Schools Initiative which

> encourages education systems to recognize happiness as both a means to and a goal of quality learning by positioning happiness as a key lever for improving learning experiences and outcomes. The Happy Schools initiative does not define or measure happiness at the individual level. Rather, it aims to improve the conditions – physical, social, pedagogical and professional – of schools to ensure they are spaces where all can flourish, experience daily joy, grow in knowledge, and practise empathy and resilience. (UNESCO, 2024, p. 10)

The same recent publication also recognizes the need for empirical support for the features of the Happy Schools Framework. Beyond the Happy Schools initiative, there is widespread interest in the state of wellbeing and school climate as reported by the members of those communities. Not only are other international initiatives such as the SDGs accounting for non-academic outcomes in their ambitions, but students, teachers, and parents are aware at the local level of the negative trend in wellbeing at school. A review of the literature shows that these constructs are complex and value-laden rendering cross-cultural comparison even at a small scale hard to achieve. Wellbeing theory comprises at least two major paradigms that are sometimes framed as opposing interests, but more often need to be integrated to fully capture the

nature of individual wellbeing. School climate definitions are influenced by complexity and social systems, requiring breadth and depth in its many facets. Responding to all these considerations towards making systems-level evaluative claims requires complex theorization and methodologies that can retain the detail of such interwoven characteristics.

Expanding the analyses and interpretation of the TIMSS Context Questionnaire data is an opportunity to take advantage of the high-quality, internationally agreed-upon data and indicators that are produced from that project to capture such features. To support the use of TIMSS data as indicators for these various goals, this study explored the state of school-level well-being based on the collection of scales available as part of the Grade 4 TIMSS 2023 Context Questionnaires. The purpose of this research was to derive patterns from student, parent, teacher, and school context data related to school climate and well-being as measured by a selection of scales administered in the context questionnaires for TIMSS 2023. This study defined school climate health to be measured by the scores on these scales towards identifying school and country groupings based on the variation in those scores observed internationally in TIMSS 2023.

This study provides a model for combining multiple background scales assessed at different levels into an indicator of value for policymakers and stakeholders at all levels, ranging from school administrators to national policy centers. These groups increasingly understand school quality to include not only academic outcomes but also non-cognitive attributes such as those available through TIMSS background and context variables. Evaluations of school quality are moving away from putting the primary focus on academic achievement, refocusing on non-academic outcomes that schools support.

This research aimed to provide a possible pathway towards empirically supporting education system evaluation internationally. International education initiatives like the SDGs

have ambitious goals for global citizenship that will require synthesis of large-scale, multidimensional data. Current investigations of school climate or student well-being have relied on data from one stakeholder group to measure what is going on in the school. Even when those measures represent perceptions of the other stakeholders or factors that might be important to those other stakeholders, this research responds to a lack of integration of stakeholder perspectives in the empirical literature. Evaluation organizations like EvalSDGs encourage countries to embrace complex systems theory to work toward the evaluation of sustainable development (Ofir et al., 2016).

To embark on a valid evaluation endeavor, it is important to understand the benchmarks and how they relate to the many stakeholders involved. An implied assumption is that there is a shared understanding of what good education entails outside of imparting academic content. Towards supporting such endeavors, this study integrated numerous scales and items administered to students, teachers, principals, and parents in 58 countries to identify multidimensional profiles of school climate health. While countries participating in TIMSS do not match the breadth of countries striving for development targets of the SDGs, it is a substantial and geographically diverse pool of countries with a strong methodological basis of international agreement. A point of clarification is necessary in that the four school clusters and three country clusters are not intended to act as benchmarks for evaluation. Instead, the hope is to spur further theorization that supports systems thinking, complexity, and integration of numerous stakeholder points of view using rigorous empirical methods.

Discussion of Results

The purpose of this dissertation was to propose a responsive methodology application for expanding the use of TIMSS data to operationalize the Happy Schools framework in nationally representative data. There were three research questions pursued in the course of this:

1. To what extent do TIMSS 2023 Grade 4 Context Questionnaire scales map to internationally established constructs of school climate and student well-being?
2. What patterns can be derived internationally from the selected response data?
3. What qualities do schools share within school climate health groups, what distinguishes the members of one group from others, and how does group membership relate to mathematics achievement?

This dissertation pursued an integrative methodology centered around multilevel latent class analysis of the school climate health scales aggregated to the school level. Response patterns on the variables were used to calculate posterior membership probabilities, determining via a modal approach to which cluster each school would be assigned. The procedure allowed for data to be used from multiple respondents, across all countries included in the sample. Schools were the Level-1 unit of analysis, allowing for the specification of groups for countries based on the proportion of schools per country assigned to each cluster. The final model selected for further analysis was four school-level clusters and three country-level classes allowing for definition of what might unify the cluster and classes.

This two-level model was the basis for all further interpretation and analysis. Defining the clusters was based on the 22 indicator variables selected from the TIMSS 2023 Context Questionnaires, based on their proximity to the variables considered in the Happy Schools and related frameworks. Each of the four clusters could be understood based on the likely values of

each variable that schools in that cluster took on. Some variables contributed little to the separation of the clusters whereas others created distinct levels when comparing clusters. Ultimately, the collection of indicators not only proved satisfactorily discriminating but resulted in cluster characterizations resonant with the realities of schools internationally. Groups were not named for country classes, but the presence of geographic and cultural similarities within groups was explored. Additionally, the proportions of clusters represented within each country was examined to describe the similarities of countries within classes in relation to the clusters of schools.

Results and implications are revisited for each of the research questions for this study.

RQ1 – TIMSS Coverage of Happy Schools

The Happy Schools Initiative and its original and updated Frameworks have broad international ambitions to shift the paradigm of international education quality toward emphasizing the emotional, psychological, and social conditions of schooling. TIMSS has a longstanding reputation for quality international comparative data bolstered by high standards for data quality and country participation. The extent that the criteria of the Happy Schools Framework could be mapped to the TIMSS 2023 Context Questionnaires to define a set of internationally valued facets of school climate and well-being captured by school climate health was a crucial first step to this research. Three phases defined this process: synthesizing Happy Schools report text into working definitions; isolating criteria that could be measured by an international comparison study like TIMSS; and aligning those criteria with items and scales from the TIMSS 2023 Grade 4 Context Questionnaires.

Each of these steps was successful in isolating the intersection of interest for this research, and also in clarifying the school climate health construct. Overall, nine criteria from the

Happy Schools framework were excluded. In the first round of exclusions, four criteria that described physical infrastructure, sanitation, and availability of activities were removed from consideration. Each of these is crucial to safe, productive schools and is perhaps particularly important to countries most targeted by initiatives like the SDGs. However, the theoretical basis for school climate health is the perceptions of community actions and attitudes relating to relationships, environment, or stakeholder satisfaction. Each of the four criteria removed at this stage focused instead on a demographic or structural component of school functioning. Thus, removing these from the definition of school climate health should not depress their importance at large. Further, demographic and descriptive factors were considered in the secondary analyses of this study as well based on what was available in the TIMSS 2023 School Questionnaire.

The alignment phase resulted in five additional criteria that were not included. Similar to the previously discussed Happy Schools criteria, these five are still viable components of a thriving school community. However, since TIMSS questionnaire items were designed independently, and not specifically constructed to measure the Happy School framework, it is sensible that there be dimensions that do not overlap. Additionally, some of the criteria are complex such that multiple TIMSS questionnaire items from multiple respondents target different pieces of a single Happy School criterion. The working definitions established in the first part of the alignment exercise were especially important during this phase. For instance, the Happy Schools criterion "Positive Discipline" at surface level seems well-aligned with several scales in TIMSS relating to disorderly conduct. However, reviewing the Happy Schools Framework documentation, this criterion was more focused on the extent of favoritism and treatment of students by adults in the school. Thus, scales in TIMSS measuring perceptions of disorderly student behavior were not as well aligned. Such mismatch could be a valuable

resource for both informing the expansion of TIMSS questionnaire items and guidance for measuring the Happy Schools framework.

The main result of this phase is the simultaneous selection of relevant Happy Schools criteria and TIMSS 2023 Context Questionnaire items that would be used in the investigation of school climate health patterns in schools internationally. The final collection represents 22 scales or items from TIMSS 2023 and 13 of 22 Happy Schools criteria. The TIMSS items and the Happy Schools Framework are both represented once more below based on the thematic grouping resulting from the literature review.

RQ2 – International Patterns of School Climate Health

Patterns of school climate health internationally were explored in their complexity through the primary methodology of this research – multilevel latent class analysis. This analysis extracted four school-level clusters and three country-level clusters to be the most fitting and meaningful model for the indicators chosen from the TIMSS 2023 Grade 4 Context Questionnaires. This exercise was successful in defining 4 distinguishable school-level clusters and 3 country-level classes based on the proportion of the four clusters present in each country. The indicator variables were discriminating overall, with the clustering explaining the most variance in the aggregated student scores. These variables were associated with Relationships and Stakeholder Satisfaction in the alignment exercise to the Happy Schools framework. This result underscores the prevalence in the literature of the role that positive interactions can have in sustaining a social environment, as they most distinguish the types of environments that exist internationally. Self-evaluation of attitudes such as students liking mathematics (ASBGSLM) or feeling confident (ASBGSCM) were also well-discriminating, echoing that student self-evaluations are responsive to school environments.

At the school level, the four clusters were characterized by the response patterns and their relative performance compared to the other clusters. The four clusters resulting from the analysis were (in the order specified):

- Cluster 1 – *System Misalignment*; orderly, safe schools with little bullying where students have trouble understanding their teachers who are dissatisfied with their jobs as are the parents in the school overall.
- Cluster 2 – *Experiencing Difficulties*; disruption and disorder reported by students, teachers, and principals while students feel little connection to their school or positive feeling towards learning mathematics or science.
- Cluster 3 – *Healthy Social Systems*; all respondent categories reporting positively on indicator variables
- Cluster 4 – *Generational Incongruence*; teachers, parents, and principals are most satisfied of any cluster, while students are more moderate, especially in their evaluations of attitudes toward mathematics and science.

Of potential interest to policy-makers and education professionals are the characteristics of Clusters 1 and 4. Clusters 2 and 3 are more easily digested as their characterization is either positive school climate health in all categories (Cluster 3) or the other end of the spectrum (Cluster 2). Clusters 1 and 4 present interesting complications in that they each contain some elements where they perform relatively poorly and some where they perform relatively well. *System Misalignment* (Cluster 2) schools are desirably stable in the sense that their environments are orderly and class time is not disrupted, but parents, students, and teachers are not satisfied with their schools. *Generational Incongruence* (Cluster 4) schools have the highest average scores from the adult stakeholders, but students are more toned down in their sense of belonging

and attitudes towards learning. These characterizations demonstrate that school climate health variables do not evenly fluctuate across school contexts – the realities are more complex. This research makes no claims of what state is satisfactory, but with so many schools outside of the *Healthy Social Systems* (Cluster 3) cluster more work is needed to interpret these groupings into workable benchmarks.

At the country level, cultural and geographical similarities were noticeable with some exceptions. Within each class, it was possible to establish at least one internal grouping such as English-speaking countries belonging to Class 3 and Asian countries belonging to Class 1. This was not the methodological grouping mechanism, which was based on the shared proportion of school-level clusters within countries. Within each cluster, there is variation in the proportion of school clusters per country that is worth further investigation and input from country experts. What the analysis demonstrates in each of the country classes overall, was that the country classes were defined by which two clusters most of their schools were assigned. Countries in Class 1 had most of their schools in either *System Misalignment* (1) or *Experiencing Difficulties* (2) clusters and had few in the *Healthy Social System* cluster (3). While countries in this group cover a broad range of average achievement, it is notable that of the highest-ranking countries in Grade 4 mathematics achievement in TIMSS 2019 that participated in TIMSS 2023, eight of them are in this group. Schools in the most prevalent clusters in these countries are not indicative of productive learning environments, on average.

Also interesting is that Cluster 1 schools, on average, mostly occur in Class 1 countries overall whereas the other three clusters have sizable proportions across the other classes. Class 2 had the most schools in the *Healthy Social System* (3) cluster, but many countries also had noticeable proportions of their schools in either Cluster 2 or 4. Considering these two classes

recalls the role of cultural differences in response patterns and in self-evaluations of constructs like wellbeing or satisfaction. Class 3 had the greatest proportions of its schools in Cluster 2 or 4, although not all countries were similar in which cluster held the majority. Uniquely striking in either case is the divide in school quality that the composition of Class 3 suggests. In these countries, students are either in schools that the adults believe are top-notch or they are in schools that are struggling to maintain a functioning social system that can support them. Students in Cluster 4 at least have more positive evaluations of their relationships and role at school, while students in Cluster 2 along with the adults have overall negative reports of their schooling. Most countries in this Class had much lower than 10% of their schools in the *Healthy Social System* (3) Cluster. Understanding how these numerous stakeholder points of view appear together in reports of the same schools has illuminated these patterns in a new, valuable manner. Further investigation is warranted into the curricular or structural identities of school systems grouped together to enrich the utility of these groupings.

RQ3 – Characteristics of School Climate Health Clusters and Achievement

Evaluating the relationship between school climate health and average school achievement could increase appeal to stakeholders. Academic attainment is an important indicator in existing national accountability programs, and TIMSS achievement measures are valuable data for these programs in particular (Kelly et al., 2020). The classifications developed in this research simultaneously describe the complexities of school climate. If shown to correlate with achievement, such evidence could also support multidimensional interventions to improve those environments as a mechanism to improve achievement.

Results of the multinomial logistic regression suggest that certain characteristics outside of school climate health are shared among the groupings. Cluster 2 (Experiencing Difficulties)

was the reference group such that comparisons to other clusters add definition to the schools most likely to belong to that group. For all clusters, schools with more linguistically diverse student populations (less than 90% speaking the language of test at home) are less likely to be in any cluster but Cluster 2 compared to schools with more than 90% of students speaking the language of test at home. This is also true for comparisons of school size and proportion of students from disadvantaged backgrounds. Schools with more than 50 Grade 4 students are less likely to be in any of the other clusters compared to Cluster 2, and schools with more than 10% of students from disadvantaged backgrounds are less likely to be in any other cluster compared to Cluster 2. This analysis is not causal such that implying the conditions cause the low climate health or the inverse are not appropriate. However, these results add to the overall understanding of how school climate health operates in schools internationally. Schools that are less linguistically homogenous and more socioeconomically diverse are less likely to experience a high-functioning school social system.

Further to this understanding, an investigation into the relationship between school climate health clusters and mathematics achievement was conducted. The findings here build on the understanding of what learning may be supported by the environments characterized by the four school climate health clusters. Perhaps not surprisingly, Cluster 2 was found to have the lowest predicted achievement with Cluster 1 predicted to have 82 points higher mean school achievement, 49 points for Cluster 3, and 41 points for Cluster 4. Along with these main effects for cluster membership, the achievement analysis also illuminated that not all students have the same benefit (as measured by school mean mathematics achievement) of different school climates. Interaction terms of school climate health cluster and the demographic variables showed for Cluster 1, 3, and 4 that schools with more linguistically diverse students and more

disadvantaged students see a reduction in the predicted achievement advantage to being in a school climate health cluster other than Cluster 2. Cluster 1 and 3 are the most striking comparisons. In the former, schools with less than 90% of students speaking the language of test at home have a predicted reduction in school mean achievement of 55 points; in Cluster 3 the reduction is 39 points, which leaves just a 10-point advantage to being in Cluster 3 compared to Cluster 2. This comparison is particularly meaningful as these two clusters represent the relative extremes of school climate health – a Healthy Social System (Cluster 3) compared to one Experiencing Difficulties (Cluster 2).

Given the known variation across countries in achievement, the final portion of this analysis was to look at the actual school mean achievement of clusters by country class. This comparison demonstrated that there were differences in achievement by cluster depending on the country class those clusters were found in. While Cluster 1 schools had roughly the same achievement across country classes, Cluster 2 achievement varied by as much as 40 points (Class 1 to Class 3) and Cluster 3 by 25 points (Class 1 to Class 2). Considering one of the ambitions of this dissertation was to support national, systems-level relationships between school climate health and achievement, a final table was presented with achievement by cluster for all countries.

These results should resonate most with policymakers as they make clear the comparative impact of school climate health on actual mean school achievement for all countries participating in TIMSS. In 50 of the 58 countries included in this analysis, schools in the *Experiencing Difficulties* (2) cluster have the lowest mean achievement. For countries interested in improving their mathematics achievement, this information is critical and previously unexplored with such a multidimensional framework simultaneously synthesizing the views of numerous stakeholders to schooling. The potential for improvement is also evidenced by this result – in nearly half of all

countries the *Healthy Social Systems* (3) schools have the highest achievement. Notably, this is the case in Korea, Singapore, and Hong Kong along with other high-performing national entities. However, consideration and caution are warranted – results need be understood within the context of the proportion of schools within each cluster per country, and with understanding of how cluster membership depends on intractable contextual features which schools cannot influence. While the socioemotional functioning of schools is important to sustain without considering the impact of achievement, the relationship is made clear by these results.

Implications and Conclusions

The results of this dissertation are revelatory if not surprising. Response patterns to the chosen scales and items from the TIMSS 2023 Grade 4 Context Questionnaire distinctly separated schools into practically meaningful typologies of school climate health. Furthermore, the proportions of these clusters grouped countries into geographically and structurally reasonable groups. Schools struggling with school climate health were most likely to have disadvantaged students and students speaking languages other than that of instruction. Within countries, it was evident that schools struggling with school climate health had the lowest mean achievement at the school level. For the highest-performing countries, the *Healthy Social Systems* (3) schools had the highest achievement. There is evidence to support the potential impact of improved school environments. However, implementation needs to be responsive to cultural specificity. It should be self-evident that universal interventions in school climate health would *not* inherently result in better schooling. Instead, national entities – or even more specific entities such as cities and towns – must reflect on what would improve the conditions summarized by school climate health. While the perceptions sourced from the Context

Questionnaires are internationally agreed upon, the qualities of the relationships that sustain school climate health and what is most valued must be responsive to the local context.

Finding that distinctions can be made in the social environments of the world's schools is a substantial step in providing descriptive clarity of what characteristics operate together within schools and where schools of different types can be found. School climate health is not exhaustive of these characteristics, nor are the secondary analyses reported here, but the scale and multidimensionality of this study are novel and have the potential to spur continued investigation. Where some variables were found to be influential, intervention can be targeted to support those schools where community members are not flourishing. Continuing the purpose of comparative international education, what can be learned from countries with higher proportions of schools in good school climate health? This study was exploratory; thus, it stands as the first step hoping to influence further theorization and implementation rather than point to 'best practices' to follow. Nonetheless, the results of the MLCA here are clear that some schools internationally are struggling to maintain a productive school environment – and they exist in almost all countries.

Further, the secondary analyses of this study demonstrate what was previously known with expansive clarity – students in social minorities are not as likely to attend schools high in school climate health. The literature summarized in this study was indicative that disadvantaged students and students speaking different languages at home than at school also experience school environments differently. The results of the secondary analyses demonstrate that these phenomena are present in the school climate health clusters, as well. However, these student experiences should be explored with the contextual expertise of local school experts and interventions informed by those additional studies. These students are from a global sample that

may share experiences overall but require unique, specific support based on other factors of life not captured in this study. Intervention is possible – wellbeing can be taught in schools as a life skill; teachers can create environments where all their students are supported to thrive. As much as this study points out that struggling schools exist in all countries, supportive schools, where students enjoy school, and teachers engage productively with their classrooms also exist in almost every country.

This study does not make claims on what is a satisfactory state, or sufficient to support the outcomes that have been defined herein. The results of this study also challenge the ambitions schooling can set out to accomplish. In many countries, it was true that the highest achievement was in clusters other than *Healthy Social System* (3), but in about half this is exactly where the highest achieving schools were. Furthermore, for many of the highest-performing countries, schools in Cluster 3 had the highest mean school achievement. Additionally, in most schools in the *Experiencing Difficulties* (2) cluster had the lowest predicted achievement. The results make clear the possibilities of supporting a healthy school environment. Thus, in considering the implications there is also a challenge set before the international education community, there seems little justification to not commit to the school climate health of schools.

Development for TIMSS 2027 could be influenced by the results of the concordance exercise between the TIMSS 2023 context questionnaires and the Happy Schools Framework. Alignment to an external international framework supports continued inclusion for those items and scales that were found to align. Furthermore, the timing is workable the expansion of measurement in TIMSS 2027 is possible by editing or adding items and scales in response. As mentioned earlier in this chapter, there were criteria in the Happy Schools Framework only

partially matched by TIMSS items – this makes for opportunity to expand what school climate health can measure and compare to the minor changes of the updated framework.

One theoretical grounding of this research was that students do not bear responsibility for the environment of their schooling. In previous studies that focused on student responses alone, this implication limits the possibilities for intervention as improvement in student perceptions cannot be directly instilled. Insofar as intervention is possible, it is limited by the systemic nature if there is no large-scale policy formulation. Thus, while some teacher and principal variables proved to discriminate little across clusters, those that did could have potential for future intervention. Finally, at risk of distracting from the novel methodology and integrative categorization of thousands of schools globally, a major takeaway is that achievement does vary across school climate health groups. Namely, the highest achieving countries have high proportions of schools that are struggling to sustain positive environments for their students.

Limitations

Identifying the boundaries around an argument strengthens the logic within by identifying what is not under consideration – "the boundary of concern" (Ulrich, 2003). The proposed research defines a multidimensional framework for school climate health such that the conditions acting on the wellbeing of school social systems can be understood internationally. Developing an understanding of wellbeing and school climate research was crucial to the comprehensiveness of this project, but the specific cognitive processes related to human mood and personality are not the intended target. Instead, the human functions that create conditions supportive of beneficial learning environments are the focus.

The study leverages the longstanding reputation of the TIMSS Context Questionnaires to make claims of international comparability. Representatives from participating countries have repeatedly approved administering them because of their importance to evaluating education environments worldwide. Nonetheless, this study is limited in that there was no possibility of influencing the content of these questionnaires to improve coverage of relevant school climate or wellbeing measures. There may be additions to future administrations after this study. An additional consideration is that the Context Questionnaires are designed to target constructs related to academic achievement, so the definition of constructs therein is construed to meet that requirement. In other words, the scales are not designed to be instruments to measure adult or child wellbeing comprehensively. The exercise to adapt the Happy School Framework demonstrates that these scales are nonetheless suitable for adaptation to established international ambitions for better schooling environments.

Furthermore, the definition of the indicator variables in the MLCA results in distributions that are group-referenced. The ordinal indicators are not norm-referenced such that values coded as 0 should not to be interpreted as 'failing' and 2 as 'excelling'. Such an inference would require significant expansion in scope including greater expertise on each domain measured than is outlined here. Instead, this exploratory study was designed to leverage an extensive international dataset toward describing the environments of schooling worldwide based on the distributions of the indicators in the TIMSS 2023 Context Questionnaire. This is a novel application of the MLCA methodology in combining data from different respondents, and in integrating the background scales reported in the TIMSS study. However, this study does not make claims that schools ought to mimic any of the characterizations described here. These

results open the door to further study of the school climate health picture painted by data gathered from stakeholders in schooling systems globally.

Insofar as TIMSS represents coherence across cultures, the nature of such a project limits the specificity of practices in education. The pursuit of comparability can omit relevant features of education systems that operate within a cultural context but not across. For instance, the Happy School Framework and related literature emphasize the positive impact that extracurricular or intramural activities can have in improving happiness. Play, personal expression, arts, or sports operate differently in relation to school across contexts. The extremes of these differences are easy to envision – in some settings, these activities *only* occur in school in some systems while *never* occurring in school in others. The contradiction results in such facets untouched by surveys like TIMSS but contributing to the overall health of students and school life.

Future Directions

This exploratory study supports further study in a multitude of directions. Within the operations of TIMSS, an accessible next step would be to engage the Grade 8 versions of the scales and data to investigate the nature of school climate health in older students. Students in secondary school are repeatedly shown to report enjoying school less, be unhappier (Gramaxo et al., 2023), and report lower levels of overall wellbeing (Tobia et al., 2019; Konu and Lintonen, 2006). The present study could essentially be replicated, but comparative research questions may also be of interest. For example, does the Grade 8 data have the same cluster solution; is the relationship with achievement comparably strong; are the clusters distinguishable, or is there less variability? The Progress in International Reading Literacy Study (PIRLS) also administers

many similar context variables; attitudes towards reading and the predictive power of school climate health of reading achievement could be interrogated.

There has been continued development in the Happy Schools Framework such that investigating the alignment between this study and those advancements may be suitable. Further, following the recently updated framework, the Happy Schools Framework has had greater exposure such that more national-level studies have been conducted. This study presents an empirically supported framework that could be well-suited for validation and adaptation within national contexts. It is within the call-to-action of the Happy Schools project that its framework be adapted to the local context. School climate health also can be the source of methodological or theoretical grounding for localization. Systems-level evaluations of school quality could leverage the holistic framing of school climate health and enhance its utility with the specification of local knowledge. This context-specific knowledge could also improve the definition of school climate health as an international measure. Those directly involved in educating students have valuable insight into the mechanisms at work in school social systems. The methodological structure of school climate health would easily integrate additional school environment characteristics that influence the atmosphere of a productive school.

References

Adler, A., & Seligman, M. E. P. (2016). Using wellbeing for public policy: Theory, measurement, and recommendations. *International Journal of Wellbeing*, *6*(1), Article 1. https://doi.org/10.5502/ijw.v6i1.429

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

Allen, K.-A., Slaten, C., Arslan, G., Roffey, S., Craig, H., & Vella-Brodrick, D. (2021). *School Belonging: The Importance of Student and Teacher Relationships* (pp. 525–550). https://doi.org/10.1007/978-3-030-64537-3_21

Allen, K.-A., Boyle, C., & Roffey, S. (2019). Creating a culture of belonging in a school context. *Educational and Child Psychology*, *36*, 5.

Alkin, M. C., Vo, A. T., & Christie, C. A. (2012). The evaluator's role in valuing: Who and with whom. *New Directions for Evaluation*, *2012*(133), 29–41. https://doi.org/10.1002/ev.20004

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

Alves, M. A., Palmer, S., & Gouveia, M. J. (2023). Psychometric Properties of the Portuguese Version of The PERMA-Profiler. *Trends in Psychology*. https://doi.org/10.1007/s43076-023-00261-3

Amholt, T. T., Dammeyer, J., Carter, R., & Niclasen, J. (2020). Psychological Well-Being and Academic Achievement among School-Aged Children: A Systematic Review. *Child Indicators Research*, *13*(5), 1523–1548. https://doi.org/10.1007/s12187-020-09725-9

Antoniou, F., Alghamdi, M. H., & Kawai, K. (2024). The effect of school size and class size on school preparedness. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1354072

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

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

Bakk, Z., & Kuha, J. (2018). Two-Step Estimation of Models Between Latent Classes and External Variables. *Psychometrika*, *83*(4), 871–892. https://doi.org/10.1007/s11336-017-9592-7

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

Berkowitz, R., Moore, H., Astor, R. A., & Benbenishty, R. (2017). A Research Synthesis of the Associations Between Socioeconomic Background, Inequality, School Climate, and Academic Achievement. *Review of Educational Research*, *87*(2), 425–469. https://doi.org/10.3102/0034654316669821

Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology: Research and applications, 2nd ed* (pp. xxii, 588). Cambridge University Press.

Borgonovi, F., & Pál, J. (2016). *A Framework for the Analysis of Student Wellbeing in the PISA 2015 Study: Being 15 In 2015*. OECD. https://doi.org/10.1787/5jlpszwghvvb-en

Bradburn, N. M. (1969). *The structure of psychological well-being.* Aldine.

Brickman, P., & Campbell, D. T. (1971). Hedonic relativism and planning the good society. In M. H. Appley (Ed.), *Adaptation level theory: A symposium* (pp. 287–302). New York: Academic Press.

Bronfenbrenner, U. (1989). Ecological systems theory. *Annals of Child Development, 6*, 187–249.

Bronfenbrenner, U. (1992). Ecological systems theory. In *Six theories of child development: Revised formulations and current issues* (pp. 187–249). Jessica Kingsley Publishers.

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

Bryk, A. S., & Schneider, B. (2002). *Trust in Schools: A Core Resource for Improvement*. Russell Sage Foundation. http://www.jstor.org/stable/10.7758/9781610440967

Butler, J., & Kern, M. L. (2016). The PERMA-Profiler: A brief multidimensional measure of flourishing. *International Journal of Wellbeing*, *6*(3), Article 3. https://doi.org/10.5502/ijw.v6i3.526

Calderon Jr, R., Kim, G., Ratsameemonthon, C., & Pupanead, S. (2020). Assessing the Adaptation of a Thai Version of the Ryff Scales of Psychological Well-Being: A PLS-SEM Approach. *Psychology*, *11*, 1037–1053. https://doi.org/10.4236/psych.2020.117068

Campbell, A., Converse, P. E., & Rodgers, W. L. (1976). *The quality of American life: Perceptions, evaluations, and satisfactions*. Russell Sage Foundation.

Cantril, H. (1965). *The pattern of human concerns*. New Brunswick, NJ: Rutgers University Press.

Care, E., Anderson, K., & Kim, H. (2016). *Visualizing the Breadth of Skills Movement Across Education Systems*. Brookings Institute. https://www.brookings.edu/articles/visualizing-the-breadth-of-skills-movement-across-education-systems/

Care, E., Kim, H., Vista, A., & Anderson, K. (2018). Education System Alignment for 21st Century Skills: Focus on Assessment. In *Center for Universal Education at The Brookings Institution*. Center for Universal Education at The Brookings Institution. https://eric.ed.gov/?id=ED592779

Cefai, C., & Cavioni, V. (2015). Beyond PISA: Schools as Contexts for the Promotion of Children's Mental Health and Wellbeing. *Contemporary School Psychology*, *19*(4), 233–242. https://doi.org/10.1007/s40688-015-0065-7

Chelimsky, E. (2006). The Purposes of Evaluation in a Democratic Society. In I. Shaw, J. Greene, & M. Mark, *The SAGE Handbook of Evaluation* (pp. 34–55). SAGE Publications Ltd. https://doi.org/10.4135/9781848608078.n1

Chandir, H. (2022). Student responses on the survey of global competence in PISA 2018. *Discourse: Studies in the Cultural Politics of Education*, *43*(4), 526–542. https://doi.org/10.1080/01596306.2020.1844153

Chandir, H., & Gorur, R. (2021). Unsustainable measures? Assessing global competence in PISA 2018. *Education Policy Analysis Archives*, *29*(August-December), Article August-December. https://doi.org/10.14507/epaa.29.4716

Chan, T. & Goldthorpe, J.. (2004). Is There a Status Order in Contemporary Britain? Evidence From the Occupational Structure of Friendship. *European Sociological Review - EUR SOCIOL REV*. 20. 383-401. 10.1093/esr/jch033.

Chang, I. (2023). Early numeracy and literacy skills and their influences on fourth-grade mathematics achievement: A moderated mediation model. *Large-Scale Assessments in Education*, *11*(1), NA-NA. https://doi.org/10.1186/s40536-023-00168-6

Chue, K. L., Yeo, A., Nie, Y., & Chew, L. C. (2023). Modifying the PERMA profiler to assess student well-being. *Current Psychology*. https://doi.org/10.1007/s12144-023-04550-z

Convention on the rights of the child (1989) Treaty no. 27531. United Nations Treaty Series, 1577, pp. 3-178. Available at:

https://treaties.un.org/doc/Treaties/1990/09/19900902%2003-14%20AM/Ch_IV_11p.pdf

Clarke, T. (2020). Children's wellbeing and their academic achievement: The dangerous discourse of 'trade-offs' in education. *Theory and Research in Education*, *18*(3), 263–294. https://doi.org/10.1177/1477878520980197

Council Resolution on a Strategic Framework for European Cooperation in Education and Training towards the European Education Area and beyond (2021-2030) 2021/C 66/01 (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32021G0226(01)

Crous, G., Casas, F., & González-Carrasco, M. (2018). What Aspects are Important to Adolescents to Achieve Full Satisfaction in Life? *Child Indicators Research*, *11*(6), 1699–1718. https://doi.org/10.1007/s12187-018-9535-6

Davidson, E. J. (2014). Evaluative reasoning. *Methodological briefs: Impact evaluation*, *4*, 1-14.

de Carvalho, T. F., de Aquino, S. D., & Natividade, J. C. (2023). Flourishing in the Brazilian context: Evidence of the validity of the PERMA-profiler scale. *Current Psychology*, *42*(3), 1828–1840. https://doi.org/10.1007/s12144-021-01587-w

Diener, E. (1984). Subjective wellbeing. *Psychological Bulletin*, *95*(3), 542–575. https://doi.org/10.1037/0033-2909.95.3.542

Diener, E. (2000). Subjective wellbeing: 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

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

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

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective wellbeing: Three decades of progress. *Psychological Bulletin*, *125*(2), 276–302. https://doi.org/10.1037/0033-2909.125.2.276

Diener, E., & Tay, L. (2015). Subjective well-being and human welfare around the world as reflected in the Gallup World Poll. *International Journal of Psychology: Journal International De Psychologie*, *50*(2), 135–149. https://doi.org/10.1002/ijop.12136

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., & Biswas-Diener, R. (2010). New Well-being Measures: Short Scales to Assess Flourishing and Positive and Negative Feelings. *Social Indicators Research*, *97*(2), 143–156. https://doi.org/10.1007/s11205-009-9493-y

Dirzyte, A., Perminas, A., & Biliuniene, E. (2021). Psychometric Properties of Satisfaction with Life Scale (SWLS) and Psychological Capital Questionnaire (PCQ-24) in the Lithuanian Population. *International Journal of Environmental Research and Public Health*, *18*. https://doi.org/10.3390/ijerph18052608

Doek, J. E. (2014). Child Wellbeing: Children's Rights Perspective. In A. Ben-Arieh, F. Casas, I. Frønes, & J. E. Korbin (Eds.), *Handbook of Child Wellbeing: Theories, Methods and Policies in Global Perspective* (pp. 187–217). Springer Netherlands. https://doi.org/10.1007/978-90-481-9063-8_9

Dreer, B. (2024). Teachers' well-being and job satisfaction: The important role of positive emotions in the workplace. *Educational Studies*, *50*(1), 61–77. https://doi.org/10.1080/03055698.2021.1940872

Etim, J. S., Etim, A., & Blizard, Z. (2022). Economically Disadvantaged, Incoming Readiness and School Achievement: Implications for Building High-Performing and Effective Schools. *Education Sciences*, *12*(8), 558. https://doi.org/10.3390/educsci12080558

Fagginger Auer, M. F., Hickendorff, M., Van Putten, C. M., Béguin, A. A., & Heiser, W. J. (2016). Multilevel Latent Class Analysis for Large-Scale Educational Assessment Data: Exploring the Relation Between the Curriculum and Students' Mathematical Strategies. *Applied Measurement in Education*, *29*(2), 144–159. https://doi.org/10.1080/08957347.2016.1138959

Fitzpatrick, J. L. (2012). An introduction to context and its role in evaluation practice. *New Directions for Evaluation*, *2012*(135), 7–24. https://doi.org/10.1002/ev.20024

Forgeard, M. J. C., Jayawickreme, E., Kern, M. L., & Seligman, M. E. P. (2011). Doing the right thing: Measuring wellbeing for public policy. *International Journal of Wellbeing*, *1*(1), Article 1. https://www.internationaljournalofwellbeing.org/index.php/ijow/article/view/15

Gagné, M., Guhn, M., Janus, M., Georgiades, K., Emerson, S. D., Milbrath, C., Duku, E., Magee, C., Schonert-Reichl, K. A., & Gadermann, A. M. (2021). Thriving, catching up, falling behind: Immigrant and refugee children's kindergarten competencies and later academic achievement. *Journal of Educational Psychology*, *113*(7), 1387–1404. https://doi.org/10.1037/edu0000634

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

Gálvez-Nieto, J. L., Vera-Bachmann, D., Trizano-Hermosilla, Í., Polanco-Levican, K., & Briceño-Olivera, C. (2020). Adaptation and Validation of the School Climate and School Identification Measure-Student Scale (SCASIM-St) in a Sample of Chilean Adolescents. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/article/10.3389/fpsyg.2020.01561

Gander, F., Hofmann, J., Proyer, R. T., & Ruch, W. (2020). Character Strengths – Stability, Change, and Relationships with Well-Being Changes. *Applied Research in Quality of Life*, *15*(2), 349–367. https://doi.org/10.1007/s11482-018-9690-4

Gates, E. F. (2016). Making sense of the emerging conversation in evaluation about systems thinking and complexity science. *Evaluation and Program Planning*, *59*, 62–73. https://doi.org/10.1016/j.evalprogplan.2016.08.004

Gates, E. F., Walton, M., Vidueira, P., & McNall, M. (2021). Introducing systems- and complexity-informed evaluation. *New Directions for Evaluation*, *2021*(170), 13–25. https://doi.org/10.1002/ev.20466

Gershenson, S., & Langbein, L. (2015). The Effect of Primary School Size on Academic Achievement. *Educational Evaluation and Policy Analysis*, *37*(1S), 135S-155S.

Giambona, F., & Porcu, M. (2018). School size and students' achievement. Empirical evidences from PISA survey data. *Socio-Economic Planning Sciences*, *64*, 66–77. https://doi.org/10.1016/j.seps.2017.12.007

Giangrasso, B. (2021). Psychometric properties of the PERMA-Profiler as hedonic and eudaimonic well-being measure in an Italian context. *Current Psychology*, *40*(3), 1175–1184. https://doi.org/10.1007/s12144-018-0040-3

Gohari, M. R., Cook, R. J., Dubin, J. A., & Leatherdale, S. T. (2020). Identifying patterns of alcohol use among secondary school students in Canada: A multilevel latent class analysis. *Addictive Behaviors*, *100*, 106120. https://doi.org/10.1016/j.addbeh.2019.106120

Góngora, V. C. (2014). Satisfaction With Life, Well-Being, and Meaning in Life as Protective Factors of Eating Disorder Symptoms and Body Dissatisfaction in Adolescents. *Eating Disorders*, *22*(5), 435–449. https://doi.org/10.1080/10640266.2014.931765

Gramaxo, P., Flores, I., Dutschke, G., & Barreto de Seabra Borges, F. I. (2023). What makes a school a happy school? Portuguese students' perspectives. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1267308

Gramaxo, P., Seabra, F., Abelha, M., & Dutschke, G. (2023). What Makes a School a Happy School? Parents' Perspectives. *Education Sciences*, *13*(4), Article 4. https://doi.org/10.3390/educsci13040375

Gregory, T., Dal, G. E., Brushe, M., Engelhardt, D., Luddy, S., Guhn, M., Gadermann, A., KA, S.-R., & Brinkman, S. (2021). Associations between School Readiness and Student Wellbeing: A Six-Year Follow Up Study. *Child Indicators Research*, *14*(1), 369–390. https://doi.org/10.1007/s12187-020-09760-6

Gregory, T., Engelhardt, D., Lewkowicz, A., Luddy, S., Guhn, M., Gadermann, A., Schonert-Reichl, K., & Brinkman, S. (2019). Validity of the Middle Years Development Instrument for Population Monitoring of Student Wellbeing in Australian School Children. *Child Indicators Research*, *12*(3), 873–899. https://doi.org/10.1007/s12187-018-9562-3

Griffin, P., & Care, E. (2014). *Assessment and Teaching of 21st Century Skills: Methods and Approach*. Springer Netherlands. http://ebookcentral.proquest.com/lib/bostoncollege-ebooks/detail.action?docID=1967645

Hargreaves, A., & Shirley, D. (2021). *Wellbeing in Schools: Three Forces That Will Uplift Your Students in a Volatile World*. Association for Supervision & Curriculum Development. http://ebookcentral.proquest.com/lib/bostoncollege-ebooks/detail.action?docID=6827065

Harzing, Anne-Wil. (2006). *Response Styles in Cross-National Survey Research: A 26-Country Study*. International Journal of Cross-Cultural Management. 6. 243-266. 10.1177/1470595806066332.

Hastedt, D., 2020. *SDG indicators deserve the best methodologies: IEA cautionary statement on political*, International Association for the Evaluation of Educational Achievement. Netherlands. Retrieved from https://policycommons.net/artifacts/1665863/sdg-indicators-deserve-the-best-methodologies/2397512/ on 29 Jan 2024.

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

Hofstede, G. (1980). Culture and Organizations. *International Studies of Management & Organization*, *10*(4), 15–41.

Hojabrian, H., Rezaei, A., Bigdeli, I., Najafi, M., & Mohammadifar, M. (2018a). Construction and Validation of the Human Psychological Flourishing Scale (HPFS) in Sociocultural Context of Iran. *Practice in Clinical Psychology*, *6*, 129–139. https://doi.org/10.29252/nirp.jpcp.6.2.129

Holte, A., Barry, M., Bekkhus, M., Borge, A., Bowes, L., Casas, F., Friborg, O., Grinde, B., Headey, B., Jozefiak, T., Lekhal, R., Marks, N., Muffels, R., Nes, R., Røysamb, E., Thimm, J., Torgersen, S., Trommsdorff, G., Veenhoven, R., & Zachrisson, H. (2014). *Psychology of Child Wellbeing* (pp. 555–631). https://doi.org/10.1007/978-90-481-9063-8_1

Holzer, J., Grützmacher, L., Lüftenegger, M., Prenzel, M., & Schober, B. (2024). Shedding light on relations between teacher emotions, instructional behavior, and student school well-being – Evidence from disadvantaged schools. *Learning and Instruction*, *92*, 101926. https://doi.org/10.1016/j.learninstruc.2024.101926

Hooper, M., Mullis, I. V. S., Martin, M. O., Fishbein, B. (2017). *TIMSS 2019 Context Questionnaire Framework*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

Huppert, F. A., & So, T. T. C. (2013). Flourishing Across Europe: Application of a New Conceptual Framework for Defining Wellbeing. *Social Indicators Research*, *110*(3), 837–861. https://doi.org/10.1007/s11205-011-9966-7

Hymel, S., & Katz, J. (2019). Designing Classrooms for Diversity: Fostering Social Inclusion. *Educational Psychologist*, *54*(4), 331–339. https://doi.org/10.1080/00461520.2019.1652098

International Association for the Evaluation of Educational Achievement. (2022, January 10). *TIMSS and PIRLS Data Contributes to SDG 4.A.2 Indicator On Bullying*. https://www.iea.nl/news-events/news/timss-and-pirls-data-contributes-sdg-4a2-indicator-bullying

Jayawickreme, E., Forgeard, M. J. C., & Seligman, M. E. P. (2012). The Engine of Well-Being. *Review of General Psychology*, *16*(4), 327-342. https://doi.org/10.1037/a0027990

Jenks, C. L. (2004). The well-being of social systems. *Systems Research and Behavioral Science*, *21*(3), 209–217. https://doi.org/10.1002/sres.616

Joshanloo, M. (2018). Optimal human functioning around the world: A new index of eudaimonic well-being in 166 nations. *British Journal of Psychology (London, England: 1953)*, *109*(4), 637–655. https://doi.org/10.1111/bjop.12316

Joshanloo, M., Van de Viert, E., & Jose, P. E. (2021) Four Fundamental Distinctions in Conceptions of Wellbeing Across Cultures. In M.L. Kern & M.L. Wehmeyer (Eds.). (2021). *The Palgrave Handbook of Positive Education*. Springer International Publishing. https://doi.org/10.1007/978-3-030-64537-3

Julnes, G. (2012), Developing policies to support valuing in the public interest. New Directions for Evaluation, 2012: 109-129. https://doi.org/10.1002/ev.20012

Juvonen, J., Lessard, L. M., Rastogi, R., Schacter, H. L., & Smith, D. S. (2019). Promoting Social Inclusion in Educational Settings: Challenges and Opportunities. *Educational Psychologist*, *54*(4), 250–270. https://doi.org/10.1080/00461520.2019.1655645

Kállay, É., & Rus, C. (2014). Psychometric Properties of the 44-Item Version of Ryff's Psychological Well-Being Scale. *European Journal of Psychological Assessment*, *30*(1), 15–21. https://doi.org/10.1027/1015-5759/a000163

Karaś, D., & Cieciuch, J. (2017). Polish adaptation of Carol Ryff's Psychological Well-Being Scales. *Roczniki Psychologiczne*, *20*, 837–853. https://doi.org/10.18290/rpsych.2017.20.4-4en

Kaya, M., & Erdem, C. (2021). Students' Well-Being and Academic Achievement: A Meta-Analysis Study. *Child Indicators Research*, *14*(5), 1743–1767. https://doi.org/10.1007/s12187-021-09821-4

Kelly, D.L., Centurino, V.A.S., Martin, M.O., & Mullis, I.V.S. (Eds.) (2020***).*** ***TIMSS 2019 Encyclopedia: Education Policy and Curriculum in Mathematics and Science.*** Retrieved

from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/encyclopedia/

Kennewell, E., Curtis, R. G., Maher, C., Luddy, S., & Virgara, R. (2022). The relationships between school children's wellbeing, socio-economic disadvantage and after-school activities: A cross-sectional study. *BMC Pediatrics*, *22*(1), 297. https://doi.org/10.1186/s12887-022-03322-1

Kern, M. L., Adler, A., Waters, L. E., & White, M. A. (2015). Measuring Whole-School Well-being in Students and Staff. In M. A. White & A. S. Murray (Eds.), *Evidence-Based Approaches in Positive Education: Implementing a Strategic Framework for Well-being in Schools* (pp. 65–91). Springer Netherlands. https://doi.org/10.1007/978-94-017-9667-5_4

Kern, M. L., & Wehmeyer, M. L. (Eds.). (2021). *The Palgrave Handbook of Positive Education*. Springer International Publishing. https://doi.org/10.1007/978-3-030-64537-3

Konishi, C., Wong, T. K. Y., Persram, R. J., Vargas-Madriz, L. F., & Liu, X. (2022). Reconstructing the concept of school climate. *Educational Research*, *64*(2), 159–175. https://doi.org/10.1080/00131881.2022.2056495

Kirkhart, K. E. (2010). Eyes on the Prize: Multicultural Validity and Evaluation Theory. *American Journal of Evaluation*, *31*(3), 400–413. https://doi.org/10.1177/1098214010373645

Konu, A., & Lintonen, T. (2006). School well-being in Grades 4-12. *Health Education Research*, *21*. https://doi.org/10.1093/her/cyl032

Krys, K., Haas, B. W., Igou, E. R., Kosiarczyk, A., Kocimska-Bortnowska, A., Kwiatkowska, A., Lun, V. M.-C., Maricchiolo, F., Park, J., Šolcová, I. P., Sirlopú, D., Uchida, Y., Vauclair,

C.-M., Vignoles, V. L., Zelenski, J. M., Adamovic, M., Akotia, C. S., Albert, I., Appoh, L., … Bond, M. H. (2023). Introduction to a Culturally Sensitive Measure of Well-Being: Combining Life Satisfaction and Interdependent Happiness Across 49 Different Cultures. *Journal of Happiness Studies*, *24*(2), 607–627. https://doi.org/10.1007/s10902-022-00588-1

Krys, K., Zelenski, J., Capaldi, C., Park, J., Tilburg, W., van Osch, Y., Haas, B., Bond, M., Dominguez Espinosa, A., Xing, C., Igbokwe, D., Kwiatkowska, A., Łużniak-Piecha, M., Nader, M., Rizwan, M., Zhu, Z., & Uchida, Y. (2019). Putting the "We" Into Well-being: Using Collectivism-Themed Measures of Well-Being Attenuates Well-being's Association With Individualism. *Asian Journal of Social Psychology*, *22*. https://doi.org/10.1111/ajsp.12364

LaRoche, S., Joncas, M., & Foy, P. (2020). Sample design in TIMSS 2019. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), **Methods and Procedures: TIMSS 2019 Technical Report** (pp. 3.1-3.33). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/methods/chapter-3.html

Ledger, S., Thier, M., Bailey, L., & Pitts, C. (2019). OECD's Approach to Measuring Global Competency: Powerful Voices Shaping Education. *Teachers College Record*, *121*(8), 1–40. https://doi.org/10.1177/016146811912100802

Lee, E., Reynolds, K. J., Subasic, E., Bromhead, D., Lin, H., Marinov, V., & Smithson, M. (2017). Development of a dual school climate and school identification measure–student (SCASIM-St). *Contemporary Educational Psychology*, *49*, 91–106. https://doi.org/10.1016/j.cedpsych.2017.01.003

Lee, V. E., & Loeb, S. (2000). School Size in Chicago Elementary Schools: Effects on Teachers' Attitudes and Students' Achievement. *American Educational Research Journal*, *37*(1), 3–31. https://doi.org/10.2307/1163470

Lee, J., & Stankov, L. (2023). Multilevel Latent Profile Analysis of Global Competence in PISA 2018: Small Country-Level Differences. *Journal of Cross-Cultural Psychology*, *54*(6–7), 658–683. https://doi.org/10.1177/00220221231183150

Leithwood, K., & Jantzi, D. (2009). A Review of Empirical Evidence about School Size Effects: A Policy Perspective. *Review of Educational Research*, *79*(1), 464–490.

López-Pérez, B., Zuffianò, A., & Benito-Ambrona, T. (2022). Cross-cultural differences in children's conceptualizations of happiness at school: European Journal of Developmental Psychology. *European Journal of Developmental Psychology*, *19*(1), 43–63. https://doi.org/10.1080/17405629.2020.1865142

Lombardi, E., Traficante, D., Bettoni, R., Offredi, I., Giorgetti, M., & Vernice, M. (2019). The Impact of School Climate on Wellbeing Experience and School Engagement: A Study With High-School Students. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02482

Lönngren, J., & van Poeck, K. (2021). Wicked problems: A mapping review of the literature. *International Journal of Sustainable Development & World Ecology*, *28*(6), 481–502. https://doi.org/10.1080/13504509.2020.1859415

Lucks, D., Etta, F., and Miranda, I.W. (2019). *EvalSDGs Insight 1: Evaluating the Complexity of the 17 SDGs*. EvalSDGs. Retrieved August 24, 2023, from https://evalsdgs.org/

Mahfooz, S.; Normén-Smith, J. *Going Global with the Happy Schools Framework: Supporting Holistic School Happiness to Improve Teaching, Learning and Well-Being*; UNESCO United Nations Educational, Scientific and Cultural Organization: Paris, France, 2022.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Markus, H. R., & Kitayama, S. (2010). Cultures and Selves: A Cycle of Mutual Constitution. *Perspectives on Psychological Science*, *5*(4), 420–430. https://doi.org/10.1177/1745691610375557

Marquez, J., & Long, E. (2021). A Global Decline in Adolescents' Subjective Wellbeing: A Comparative Study Exploring Patterns of Change in the Life Satisfaction of 15-Year-Old Students in 46 Countries. *Child Indicators Research*, *14*(3), 1251–1292. https://doi.org/10.1007/s12187-020-09788-8

Marsh, H., Huppert, F., Donald, J., Horwood, M., & Sahdra, B. (2019). The well-being profile (WB-Pro): Creating a theoretically based multidimensional measure of well-being to advance theory, research, policy, and practice. *Psychological Assessment*, *32*. https://doi.org/10.1037/pas0000787

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149-174.

Maxwell, S., Reynolds, K. J., Lee, E., Subasic, E., & Bromhead, D. (2017). The Impact of School Climate and School Identification on Academic Achievement: Multilevel Modeling with Student and Teacher Data. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.02069

Miller, P., & Votruba-Drzal, E. (2015). Urbanicity Moderates Associations between Family Income and Adolescent Academic Achievement. *Rural Sociology*, *80*(3), 362–386. https://doi.org/10.1111/ruso.12067

Mitchell, M. M., Bradshaw, C. P., & Leaf, P. J. (2010). Student and Teacher Perceptions of School Climate: A Multilevel Exploration of Patterns of Discrepancy. *Journal of School Health*, *80*(6), 271–279. https://doi.org/10.1111/j.1746-1561.2010.00501.x

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

Mullis, I. V. S., & Fishbein, B. (2020). Updating the TIMSS 2019 instruments for describing the contexts for student learning. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), **Methods and Procedures: TIMSS 2019 Technical Report** (pp. 2.1-2.9). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/methods/chapter-2.html

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-results/

Muthén, B. (2004). Latent Variable Analysis: Growth Mixture Modeling and Related Techniques for Longitudinal Data.

Mutz, R., & Daniel, H.-D. (2013). University and student segmentation: Multilevel latent-class analysis of students' attitudes towards research methods and statistics. *British Journal of Educational Psychology*, *83*(2), 280–304. https://doi.org/10.1111/j.2044-8279.2011.02062.x

OECD. (2015). *How's Life? 2015: Measuring Wellbeing*. OECD. https://doi.org/10.1787/how_life-2015-en

OECD. (2019a). *PISA 2018 Assessment and Analytical Framework*. OECD. https://doi.org/10.1787/b25efab8-en

OECD. (2019b). *PISA 2018 Global Competence Framework* (pp. 165–215). OECD. https://doi.org/10.1787/043fc3b0-en

OECD. (2019c). *PISA 2018 Results (Volume III): What School Life Means for Students' Lives*. OECD. https://doi.org/10.1787/acd78851-en

OECD. (2019d). *PISA 2018 Wellbeing Framework* (pp. 257–298). OECD. https://doi.org/10.1787/38a34353-en

Ofir, Z. & Schwandt, T. (2016). *Five considerations for national evaluation agendas informed by the SDGs*. International Institute for Environment and Development. Retrieved August 24, 2023, from https://www.iied.org/17374iied

Ofir, Z., & Rugg, D. (2021). *American Journal of Evaluation* Section on International Developments in Evaluation: Transforming Evaluation for Times of Global Transformation. *American Journal of Evaluation*, *42*(1), 47–52. https://doi.org/10.1177/1098214020979070

Pineda Rodriguez, P. & Sandoval-Hernandez, A., 2023. *Guidelines for Reviewing and Integrating Education Assessments into the Evaluation of SDG4 Indicators*, UNESCO

Institute for Statistics. Paris, France. https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2023/12/WG_GAML10_ST.REP_1_Assessment-blueprint.pdf

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

Ramlackhan, K., & Wang, Y. (2024). Urban School District Performance: A Longitudinal Analysis of Achievement. *Urban Education*, *59*(3), 849–877. https://doi.org/10.1177/00420859211044947

Rappleye, J., Komatsu, H., Uchida, Y., Krys, K., & Markus, H. (2020). 'Better policies for better lives'?: Constructive critique of the OECD's (mis)measure of student wellbeing. *Journal of Education Policy*, *35*(2), 258–282. https://doi.org/10.1080/02680939.2019.1576923

Ricciardi, C., Manfra, L., Hartman, S., Bleiker, C., Dineheart, L., & Winsler, A. (2021). School readiness skills at age four predict academic achievement through 5th grade. *Early Childhood Research Quarterly*, *57*, 110–120. https://doi.org/10.1016/j.ecresq.2021.05.006

Riley, K. (2019). Agency and belonging: What transformative actions can schools take to help create a sense of place and belonging? *Educational and Child Psychology*, *36*, 91–104. https://doi.org/10.53841/bpsecp.2019.36.4.91

Rosenmann, A., & Kurman, J. (2019). The culturally situated self. In *The handbook of culture and psychology, 2nd ed* (pp. 538–585). Oxford University Press. https://doi.org/10.1093/oso/9780190679743.003.0017

Rudasill, K. M., Snyder, K. E., Levinson, H., & L. Adelson, J. (2018). Systems View of School Climate: A Theoretical Framework for Research. *Educational Psychology Review*, *30*(1), 35–60. https://doi.org/10.1007/s10648-017-9401-y

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

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, *69*(4), 719–727. https://doi.org/10.1037/0022-3514.69.4.719

Sabol, T. J., & Pianta, R. C. (2012). Patterns of School Readiness Forecast Achievement and Socioemotional Development at the End of Elementary School. *Child Development*, *83*(1), 282–299.

Sälzer, C., & Roczen, N. (2018). Assessing Global Competence in PISA 2018: Challenges and Approaches to Capturing a Complex Construct. *International Journal of Development Education and Global Learning*, *10*(1), 5–20.

Schonert-Reichl, K. A., Guhn, M., Gadermann, A. M., Hymel, S., Sweiss, L., & Hertzman, C. (2013). Development and Validation of the Middle Years Development Instrument (MDI): Assessing Children's Well-Being and Assets across Multiple Contexts. *Social Indicators Research*, *114*(2), 345–369. https://doi.org/10.1007/s11205-012-0149-y

Schwandt, T., Ofir, Z., Lucks, D., El-Saddick, K., and D'Errico, S. (2016a). *Evaluation: a crucial ingredient for SDG success*. International Institute for Environment and Development. Retrieved August 24, 2023, from https://www.iied.org/17357iied

Schwandt, T., Ofir, Z., Lucks, D., El-Saddick, K., and D'Errico, S. (2016b). *Realising the SDGs by reflecting on the way(s) we reason, plan and act: the importance of evaluative thinking*. International Institute for Environment and Development. Retrieved August 24, 2023, from https://www.iied.org/17380iied

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.

Seligman, M. E. P. (2002). Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment. New York, The Free Press.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, *55*(1), 5–14. https://doi.org/10.1037/0003-066X.55.1.5

Sheldon, K. M., Titova, L., Gordeeva, T. O., Osin, E. N., Lyubomirsky, S., & Bogomaz, S. (2017). Russians inhibit the expression of happiness to strangers: Testing a display rule model. *Journal of Cross-Cultural Psychology*, *48*(5), 718–733. https://doi.org/10.1177/0022022117699883

Shirley, D. (2020). Beyond Well-being: The Quest for Wholeness and Purpose in Education. *ECNU Review of Education*, *3*(3), 542–555. https://doi.org/10.1177/2096531120932579

Stavraki, M., García-Márquez, R., Bajo, M., Callejas-Albiñana, A. I., Paredes, B., & Díaz, D. (2022). Brief Version of the Ryff Psychological Well-Being Scales for Children and Adolescents: Evidence of Validity. *Psicothema*, *34.2*, 316–322. https://doi.org/10.7334/psicothema2021.235

Strelhow, M. R. W., Sarriera, J. C., & Casas, F. (2020). Evaluation of Well-Being in Adolescence: Proposal of an Integrative Model with Hedonic and Eudemonic Aspects. *Child Indicators Research*, *13*(4), 1439–1452. https://doi.org/10.1007/s12187-019-09708-5

Suh, E., Diener, E., Oishi, S., & Triandis, H. (1998). The Shifting Basis of Life Satisfaction Judgments Across Cultures: Emotions Versus Norms. *Journal of Personality and Social Psychology*, *74*, 482–493. https://doi.org/10.1037/0022-3514.74.2.482

Suh, E. M., Diener, E., & Updegraff, J. A. (2008). From Culture To Priming Conditions: Self-Construal Influences on Life Satisfaction Judgments. *Journal of Cross-Cultural Psychology*, *39*(1), 3–15. https://doi.org/10.1177/0022022107311769

Sulak, T. N. (2016). School Climate and Academic Achievement in Suburban Schools. *Education and Urban Society*, *48*(7), 672–684. https://doi.org/10.1177/0013124514541465

Sulak, T. N. (2018). School climate: The controllable and the uncontrollable. *Educational Studies*, *44*(3), 279–294. https://doi.org/10.1080/03055698.2017.1373630

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

Tomul, E., Önder, E., & Taslidere, E. (2021). The relative effect of student, family and school-related factors on math achievement by location of the school. *Large-Scale Assessments in Education*, *9*(1), 22. https://doi.org/10.1186/s40536-021-00117-1

Tov, W. (2009). Culture and subjective wellbeing. In E. Diener (Ed.), *Culture and wellbeing: The collected works of Ed Diener* (pp. 9–41). Springer Science + Business Media. https://doi.org/10.1007/978-90-481-2352-0_2

Triandis, H. C. (1996). The psychological measurement of cultural syndromes. *American Psychologist*, *51*(4), 407–415. https://doi.org/10.1037/0003-066X.51.4.407

Transforming our world: The 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs. (n.d.). Retrieved July 17, 2023, from https://sdgs.un.org/2030agenda

Uchida, Y., & Rappleye, J. (2024). An Interdependent Approach: Manifestations in Cultural Practices. In Y. Uchida & J. Rappleye (Eds.), *An Interdependent Approach to Happiness and Well-Being* (pp. 97–128). Springer International Publishing. https://doi.org/10.1007/978-3-031-26260-9_5

Ulrich, W. (2003). Beyond methodology choice: critical systems thinking as critically systemic discourse. *Journal of the Operational Research Society*, *54*, 325-342.

UNESCO. (2024). *Why the world needs happy schools: Global report on happiness in and for learning*. Paris: UNESCO.

UNESCO (2021) *Definition of Sustainable Development Goal (SDG) thematic indicator 4.a.2: percentage of students who experienced bullying in the past 12 months*. https://unesdoc.unesco.org/ark:/48223/pf0000378924?locale=en

UNESCO and IEA. (2020). *Measuring global education goals: how TIMSS helps monitoring progress towards Sustainable Development Goal 4 using TIMSS (Trends in International Mathematics and Science Study)*. https://unesdoc.unesco.org/ark:/48223/pf0000375119

UNESCO. (2014). Learning to Live Together. Education Policies and Realities in the Asia-Pacific. Paris, UNESCO. http://unesdoc.unesco.org/images/0022/002272/227208E.pdf (Accessed 15 October 2015.)

UNESCO Bangkok. *Happy schools! A framework for learner wellbeing in the asia-pacific*. (2020, September 8). Retrieved March 30, 2023, from

https://bangkok.unesco.org/content/happy-schools-framework-learner-wellbeing-asia-pacific

UNESCO Institute of Statistics. (2022). *Monitoring of the Sustainable Development Goals Using Large-Scale International Assessments.* https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2022/04/Monitoring-of-the-SDGs-Using-Large-Scale-International-Assessments_April-2022.pdf

UNESCO Institute of Statistics. (2019). *SDG 4 Data Book GLOBAL EDUCATION INDICATORS 2019***.** https://uis.unesco.org/sites/default/files/documents/sdg4-databook-global-ed-indicators-2019-en.pdf

UNESCO Institute of Statistics. (2018a). *Quick Guide to Education Indicators for SDG 4*. https://uis.unesco.org/sites/default/files/documents/quick-guide-education-indicators-sdg4-2018-en.pdf

UNESCO Institute of Statistics. (2018b). *SDG 4 Data Book GLOBAL EDUCATION. INDICATORS 2018* https://uis.unesco.org/sites/default/files/documents/sdg4-data-book-2018-en.pdf

UNESCO and IEA. (2020). *Measuring global education goals: how TIMSS helps monitoring progress towards Sustainable Development Goal 4 using TIMSS (Trends in International Mathematics and Science Study)*. https://unesdoc.unesco.org/ark:/48223/pf0000375119

United Nations. (2015). *The 17 goals | sustainable development*. United Nations. Retrieved March 30, 2023, from https://sdgs.un.org/goals

UN General Assembly, *Declaration of the Rights of the Child*, 20 November 1959, A/RES/1386(XIV), available at: https://www.refworld.org/docid/3ae6b38e3.html [accessed 2 October 2023]

Valiente, C., Doane, L. D., Clifford, S., Grimm, K. J., & Lemery-Chalfant, K. (2021). School readiness and achievement in early elementary school: Moderation by Students' temperament. *Journal of Applied Developmental Psychology*, *74*, 101265. https://doi.org/10.1016/j.appdev.2021.101265

Van Der Wildt, A., Van Avermaet, P., & Van Houtte, M. (2017). Multilingual school population: Ensuring school belonging by tolerating multilingualism. *International Journal of Bilingual Education and Bilingualism*, *20*(7), 868–882. https://doi.org/10.1080/13670050.2015.1125846

Vella-Brodrick, D. A., Gill, A., & Patrick, K. (2022). Seeing Is Believing: Making Wellbeing More Tangible. *Frontiers in Psychology*, *13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.809108

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

Vermunt, J.K. and Magidson J. (2016). Technical Guide for Latent GOLD 5.1: Basic, Advanced, and Syntax. Belmont, MA: Statistical Innovations Inc.

Vermunt, J.K., and Magidson J. (2005). Latent GOLD 4.0 User's Guide. Belmont, Massachusetts: Statistical Innovations Inc.

Viejo, C., Gómez-López, M., & Ortega-Ruiz, R. (2018). Adolescents' Psychological Well-Being: A Multidimensional Measure. *International Journal of Environmental Research and Public Health*, *15*(10), Article 10. https://doi.org/10.3390/ijerph15102325

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

Vujčić, M. T., Brajša-Žganec, A., & Franc, R. (2019). Children and Young Peoples' Views on Well-Being: A Qualitative Study. *Child Indicators Research*, *12*(3), 791–819. https://doi.org/10.1007/s12187-018-9559-y

Wang, M.-T., & Eccles, J. (2013). School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction*, *28*, 12–23. https://doi.org/10.1016/j.learninstruc.2013.04.002

Waugh, M., & Guhn, M. (2014). Bioecological Theory of Human Development. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Wellbeing Research* (pp. 398–401). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_3438

Waxman, H. C., Rivera, H. H., Linn, B., Padrón, Y. N., Rollins, K. B., Boriack, A. W., & Alford, B. A. (2016). Development of an observation instrument to measure flourishing learning environments. *Journal of Chinese Studies*, *1*(1), 3. https://doi.org/10.1186/s40853-016-0005-9

Wei, R. C., Pecheone, R. L., & Wilczak, K. L. (2015). Measuring what really matters. *Phi Delta Kappan*, *97*(1), 8–13. https://doi.org/10.1177/0031721715602228

Wortham, S., Love-Jones, R., Peters, W., Morris, S., & García-Huidobro, J. C. (2020). Educating for Comprehensive Well-being. *ECNU Review of Education*, *3*(3), 406–436. https://doi.org/10.1177/2096531120928448

Yamaguchi, K. (2000). Multinomial Logit Latent-Class Regression Models: An Analysis of the Predictors of Gender Role Attitudes among Japanese Women. *American Journal of Sociology*, *105*(6), 1702–1740. https://doi.org/10.1086/210470

Yin, L., & Fishbein, B. (2020). Creating and interpreting the TIMSS 2019 context questionnaire scales. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), Methods and Procedures: TIMSS 2019 Technical Report (pp. 16.1-16.331). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/methods/chapter-16.html

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

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

Appendices

Appendix A: Full Item and Stem for TIMSS 2023 Analysis

Relationships

Teacher

G12	How often do you do the following in teaching this class?	a) Relate the lesson to students' daily lives b) Ask students to explain their answers c) Communicate goals or objectives for the lesson to the students d) Ask students to complete challenging exercises that require them to go beyond the instruction e) Encourage classroom discussions among students f) Link new content to students' prior knowledge g) Ask students to decide their own problem solving procedures

Student

MS4	How often do these things happen in your mathematics lessons?	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 a) f) Other students' behavior makes it hard for me to concentrate
MS9	How often do these things happen in your science lessons?	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 b) f) Other students' behavior makes
MS3	How much do you agree with these statements about your mathematics lessons?	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 h) My teacher asks me to show what I have learned i) My teacher asks me to explain my answers
MS8	How much do you agree with these statements about your science lessons?	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 h) My teacher asks me to show what I have learned i) My teacher asks me to explain my answers

Principal

14	To what degree is each of the following a problem among <fourth grade> students in your school?	a) Arriving late at school b) Absenteeism (i.e., unjustified absences) c) Classroom disturbance d) Cheating e) Profanity f) Vandalism g) Theft h) Intimidation or verbal abuse among students (including texting, emailing, etc.) i) Physical fights among students j) Intimidation or verbal abuse of teachers or staff (including texting, emailing, etc.)

Environment

Teacher

G6	How would you characterize each of the following within your school?	a) Teachers' understanding of the school's curricular goals b) Teachers' degree of success in implementing the school's curriculum c) Teachers' expectations for student achievement d) Teachers' ability to inspire students e) Parental involvement in school activities f) Parental commitment to ensure that students are ready to learn g) Parental expectations for student achievement h) Parental support for student achievement i) Students' desire to do well in school j) Students' ability to reach school's academic goals k) Students' respect for classmates who excel academically
G13	In your view, to what extent do the following limit how you teach this class?	a) Students lacking prerequisite knowledge or skills b) Students suffering from lack of basic nutrition c) Students suffering from not enough sleep d) Students absent from class e) Disruptive students f) Uninterested students g) Distracted students h) Students with mental, emotional, or psychological impairment i) Students with difficulties understanding the language of instruction
G7	How much do you agree or disagree with the following statements about your current school?	a) I feel safe at this school b) This school's security policies and practices are sufficient c) The students behave in an orderly manner d) The students are respectful of the teachers e) The students respect school property f) This school has clear rules about student conduct g) This school's rules are enforced in a fair and consistent manner

Student

G13	What do you think about your school? Tell how much you agree with these statements.	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 my school care about me e) I am proud to go to this school f) I have friends at this school

		g) Students at this school like me the way I am
G9	How often do you feel this way when you arrive at school? (hungry/tired)	a) I feel tired c) b) I feel hungry
G14	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?	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

Principal

12	How would you characterize each of the following within your school?	a) Teachers' understanding of the school's curricular goals b) Teachers' degree of success in implementing the school's curriculum c) Teachers' expectations for student achievement d) Teachers' ability to inspire students e) Parental involvement in school activities f) Parental commitment to ensure that students are ready to learn g) Parental expectations for student achievement h) Parental support for student achievement i) Students' desire to do well in school j) Students' ability to reach school's academic goals k) Students' respect for classmates who excel academically
11	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) b) Supplies (e.g., papers, pencils, materials) c) School buildings and grounds d) Heating/cooling and lighting systems e) Instructional space (e.g., classrooms) f) Technologically competent staff g) Audio-visual resources for delivery of instruction (e.g., interactive white boards, digital projectors) h) Digital devices for student use i) Resources for students with disabilities
	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 b) Computer software/applications for mathematics instruction c) Library resources relevant to mathematics instruction d) Calculators for mathematics instruction e) Concrete objects or materials to help students understand quantities or procedures
	How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following? [Resources for Science Instruction]	a) Teachers with a specialization in science b) Computer software/applications for science instruction c) Library resources relevant to science instruction d) Science equipment and materials for experiments

Stakeholder Satisfaction

Teacher

G8	How often do you have these feelings about being a teacher?	a) I am content with my profession as a teacher b) I find my work full of meaning and purpose c) I am enthusiastic about my job d) My work inspires me e) I am proud of the work I do f) I feel appreciated as a teacher g) I enjoy the challenges of teaching
G9	How much do you agree or disagree with the statements below?	a) There are too many students in the classes b) I have too much material to cover in class c) I have too many teaching hours d) I need more time to prepare for class e) I need more time to assist individual students f) I feel too much pressure from parents g) I have difficulty keeping up with all of the changes to the curriculum h) I have too many administrative tasks

Student

MS2	How much do you agree with these statements about learning mathematics?	a) I enjoy learning mathematics b) I wish I did not have to study mathematics c) Mathematics is boring 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
MS5	How much do you agree with these statements about mathematics?	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
MS7	How much do you agree with these statements about learning science?	a) I enjoy learning science b) I wish I did not have to study science c) Science is boring 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
MS10	How much do you agree with these statements about science?	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

Principal

15	To what degree is each of the following a problem among teachers in your school?	a) Arriving late or leaving early b) Absenteeism

Parent

8	What do you think of your child's school?	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 B: R Code for Dataset creation

```
Teachers <- read_sav("G:/User/Charlotte/T23data/SPSS/Teacher_Unique.sav")
Principals <- read_sav("G:/User/Charlotte/T23data/SPSS/School_Unique.sav")
Students <- read_sav("G:/User/Charlotte/T23data/SPSS/StudHome_Concat.sav")

Parents$UniqueSchool <- paste(Parents$IDCNTRY, Parents$IDSCHOOL, sep = "_")
Parents$UniqueStud <- paste(Parents$IDCNTRY, Parents$IDSTUD, sep = "_")

Principals$UniqueSchool <- paste(Principals$IDCNTRY, Principals$IDSCHOOL, sep = "_")


Students$UniqueSchool <- paste(Students$IDCNTRY, Students$IDSCHOOL, sep = "_")
Students$UniqueStud <- paste(Students$IDCNTRY, Students$IDSTUD, sep = "_")

Teachers$UniqueSchool <- paste(Teachers$IDCNTRY, Teachers$IDSCHOOL, sep = "_")
Teachers$UniqueStud <- paste(Teachers$IDCNTRY, Teachers$IDSTUD, sep = "_")

Principals <- subset(Principals, select = -UniqueStud)
data2 <- subset(data2, select = -ACBG13)
data2 <- subset(data2, select = -ACBG16)

data1<-Parents
data2<-Principals
data3<-Teachers
data4<-Students

target_vars2<-c("ACBG15")
for(var_num in 1:1){
  var<-target_vars2[var_num]
  cols_to_combine2<-names(data2)[grepl(var,names(data2))]
  data2[[var]]<-rowSums(data2[names(data2)%in%cols_to_combine2],na.rm=T)
}

target_vars3<-c("ATBG09","ATBG12")
for (var_num in 1:2){
  var3<-target_vars3[var_num]
  cols_to_combine3<-names(data3)[grepl(var3,names(data3))]
  data3[[var3]]<-rowSums(data3[names(data3)%in%cols_to_combine3],na.rm=T)
}

target_vars4<-c("ASBG09")
for (var_num in 1:1){
  var4<-target_vars4[var_num]
  cols_to_combine4<-names(data4)[grepl(var4,names(data4))]
  data4[[var4]]<-rowSums(data4[names(data4)%in%cols_to_combine4],na.rm=T)
}

Scored_Stud<-subset(data4, select=c("UniqueSchool", "ASBG09", "totwgt"))
Scored_Teach<-subset(data3, select=c("UniqueSchool", "ATBG09","ATBG12", "tchwgt"))
Scored_School<-subset(data2, select=c("UniqueSchool", "ACBG15", "SCHWGT"))
SchoolScales<-subset(data2, select=c("UniqueSchool", "ACBGMRS", "ACBGEAS","ACBGDAS"))
CountryVars<-subset(data2, select=c("UniqueSchool", "countryl", "IDCNTRY"))

write_sav(Scored_Stud,"G:/User/Charlotte/T23data/SPSS/StudScores.sav")
write_sav(Scored_Teach,"G:/User/Charlotte/T23data/SPSS/TeachScores.sav")
write_sav(Scored_School,"G:/User/Charlotte/T23data/SPSS/SchoolScorest.sav")

StudScales<-read_xlsx("C:/Users/aldricch/Desktop/Dissertation/StudHome_Scale_Mean.xlsx")
ASBG09<-read_xlsx("C:/Users/aldricch/Desktop/Dissertation/Stud_ASBG09_Mean.xlsx")
```

```
TeachScales<-read_xlsx("C:/Users/aldricch/Desktop/Dissertation/Teacher_Scale_Mean.xlsx")
ATBG0912<-read_xlsx("C:/Users/aldricch/Desktop/Dissertation/Teach_ATBG09_ATBG12_Mean.xlsx")

DissData1<-full_join(Scored_School, SchoolScales, by = "UniqueSchool")
DissData1<-full_join(DissData1, StudScales, by = "UniqueSchool")
DissData1<-full_join(DissData1, ASBG09, by = "UniqueSchool")
DissData1<-full_join(DissData1, TeachScales, by = "UniqueSchool")
DissData1<-full_join(DissData1, ATBG0912, by = "UniqueSchool")
DissData1<-full_join(DissData1, CountryVars, by = "UniqueSchool")

write_sav(DissData1,"G:/User/Charlotte/T23data/SPSS/DissData.sav")
write_xlsx(DissData1,"G:/User/Charlotte/T23data/SPSS/DissData.xlsx")
schwgt<-data2[,c("UniqueSchool", "SCHWGT")]

TestData1<-full_join(reduced2, reduced3, by = "UniqueSchool")
TestData1<-full_join(TestData1, reduced4, by = "UniqueSchool")
TestData1<-full_join(TestData1, reduced5, by = "UniqueSchool")
TestData1<-full_join(TestData1, reduced6, by = "UniqueSchool")
TestData1<-full_join(TestData1, reduced7, by = "UniqueSchool")
TestData1<-full_join(TestData1, reduced8, by = "UniqueSchool")

TestData_schwgt<-full_join(TestData1, schwgt, by = "UniqueSchool")

write_excel_csv(TestData1, "C:/Users/aldricch/Desktop/Dissertation/Chapter 3/T19/T19 All
Trial/All_trial_data.csv")
write_excel_csv(TestData_schwgt,"C:/Users/aldricch/Desktop/Dissertation/Chapter 3/T19/T19 All
Trial/All_trial_data_schwgt.csv")


##Getting Assignment to match to original data
Model_Assignment<-read_sav("G:/User/Charlotte/T23data/SPSS/assignment.sav")
DissData_redux<-read_sav("G:/User/Charlotte/T23data/SPSS/DissData_IndicatorsOnly.sav")
Assignment_redux<-subset(Model_Assignment, select=c("UniqueSchool", "countryl",
"clu#","GClass#"))
NoCode_Assignment<-full_join(DissData_redux, Assignment_redux, by = "UniqueSchool")
write_sav(NoCode_Assignment,"G:/User/Charlotte/T23data/SPSS/NoCodeWithClusterAssignment.sav")

##Setting up dataset for secondary analyses
Model_Assignment<-read_sav("G:/User/Charlotte/T23data/SPSS/assignment.sav")
Assignment_redux<-subset(Model_Assignment, select=c("UniqueSchool", "countryl",
"clu#","GClass#"))

target_vars2<-c("ACBG16")
for(var_num in 1:1){
  var<-target_vars2[var_num]
  cols_to_combine2<-names(data2)[grepl(var,names(data2))]
  data2[[var]]<-rowSums(data2[names(data2)%in%cols_to_combine2],na.rm=T)
}

Second_Stud<-subset(Students, select=c("UniqueSchool", "asmmat01",
"asmmat02","asmmat03","asmmat04", "asmmat05","totwgt"))
Second_School<-subset(data2, select=c("UniqueSchool","ACBG01","ACBG02", "ACBG03A", "ACBG03B",
"ACBG04", "ACBG05A","ACBG05B", "ACBG16", "SCHWGT"))

Covar<-full_join(Assignment_redux, Second_School, by = "UniqueSchool")
Covar_Achieve<-full_join(Covar, Second_Stud, by = "UniqueSchool")

write_sav(Covar_Achieve,"G:/User/Charlotte/T23data/SPSS/Second_with_Achievement.sav")

write_sav(Covar,"G:/User/Charlotte/T23data/SPSS/Second.sav")
```

```
write_sav(Second_Stud,"G:/User/Charlotte/T23data/SPSS/Achievement_to_Calculate.sav")


Second_Ach<-read_sav("G:/User/Charlotte/T23data/SPSS/Second.sav")
Ach<-read_xlsx("G:/User/Charlotte/T23data/SPSS/Achievement Means.xlsx")
Ach<-subset(Ach, select=c("UniqueSchool", "School_Mean_Achievement"))
Second_Ach<-full_join(Second_Ach, Ach, by = "UniqueSchool")
       write_sav(Second_Ach, "G:/User/Charlotte/T23data/SPSS/Second_with_A
```