Rethinking the contribution of school infrastructure to educational outcomes through the lens of complexity

By Darryl Walker

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Professor Catherine Killen and Doctor Leila Moslemi Naeni

University of Technology Sydney
Faculty of Design Architecture and Building

July 2025

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Darryl John Walker, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Built Environment at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

Darryl John Walker................................

7 July 2025

Abstract

This thesis examines how students' experiences of school infrastructure, conceptualised as the Institutional Environment within the School Climate framework, predict educational outcomes in secondary schools when these schools are viewed as complex, dynamic systems.

The motivation for this research stems from a fundamental tension in educational policy and research. While school infrastructure requires significant capital and operational investment, traditional Evidence-Based Policymaking typically deems it to have minimal direct impact on educational outcomes. In contrast, emerging studies suggest that infrastructure may exert important indirect influences, such as through teacher satisfaction or student attendance, prompting some researchers to reassess its role. Simultaneously, School Climate research increasingly recognises infrastructure as an integral element of the broader system shaping school experience and outcomes. This study responds to these developments by re-examining infrastructure not as a peripheral backdrop but as a potentially active driver of educational outcomes.

Grounded in pragmatism and informed by complexity theory, the research investigates context-specific, non-linear relationships between the students' experiences of infrastructure and student outcomes. Using an adapted School Climate survey in ten government-funded secondary schools in New South Wales, the study analyses student experience and student-reported wellbeing and behaviour data. X-means clustering identifies outcome groups, and gradient boosted trees (GBTs) predict cluster membership from 73 School Climate attributes, including 23 infrastructure-specific measures.

Four key findings emerge. First, student outcomes can be clustered meaningfully across all schools, preserving their multidimensional nature and reflecting complex configurations of wellbeing and behaviour rather than reducing multi-faceted outcome constructs to isolated metrics. Second, the predictive relationships between School Climate attributes and outcomes are found to be context-dependent and non-linear, highlighting the limits of generalisable, predominantly linear, models describing statistical relationships between educational outcomes and the factors believed to influence them. Third, students' perceptions of infrastructure—specifically lighting, cleanliness, spatial adequacy, and ambient comfort—consistently emerge as strong predictors of wellbeing and behaviour, positioning infrastructure as a central rather than peripheral factor in shaping student experience. Finally, the findings demonstrate that GBTs can effectively model the complexity of educational systems and potentially offer predictive insights without oversimplifying student experience.

Theoretically, this research further embeds complexity into educational discourse and validates infrastructure as a core element of School Climate. Methodologically, it demonstrates the value of clustering and GBTs in modelling system interdependencies while maintaining fidelity of student voice. This research supports the inclusion of student-reported infrastructure experiences when considering school improvement strategies and policy. The findings highlight the potential for context-sensitive, evidence-informed, and complexity-aware approaches, recognising infrastructure as an active, but typically overlooked, driver of student wellbeing and behaviour, particularly in disadvantaged contexts.

Acknowledgement

Much like the focus of this thesis, the journey from inception to completion has been defined by the interplay between simplicity and complexity. Had this been an infrastructure project, it would undoubtedly have exceeded its budget, missed every deadline, and emerged bearing little resemblance to the original plan.

Many of the obstacles were unexpected, at least to me. Most reflected my initial mindset: overconfident, under-informed, too eager to speak, and not nearly inclined enough to read, listen, and reflect. These might be forgivable traits in young professionals, students or early-career researchers, what we often excuse as the folly of youth. I, however, can make no such claim. I embarked on this path in my late forties.

What began as a whimsical pursuit came dangerously close to ending the same way on more than one occasion. This thesis exists in no small part due to the unique contributions of three extraordinary and brilliant women.

I have no doubt that my principal supervisor, Professor Cathy Killen, has sometimes questioned the wisdom of entertaining that first meeting; an encounter marked by incoherent ideas and inflated aspirations loosely wrapped in enthusiasm. Perhaps she accepted the challenge out of professional curiosity or academic goodwill. Whatever the reason, Cathy remained steadfast throughout. Her patience, clarity, and ability to impose structure on chaos were instrumental in guiding me from vague ambition to coherent contribution. Her unwavering professionalism was matched by a kindness that never diminished, no matter the obstacles.

My secondary supervisor, Dr Leila Naeni, joined the University of Technology Sydney (UTS) roughly the same time I did. Leila's mind is sharply analytical and uncompromising in its rigour. With her, there are no shortcuts, vague assertions or untested assumptions. Leila challenged my thinking with precision, firmness, and fairness. Her feedback raised the bar repeatedly, and her quiet but resolute presence reminded me to listen carefully.

Together, Cathy and Leila fundamentally reshaped how I think, engage with, and approach research. They taught me far more than I could have imagined at the outset. I can only hope that, in their eyes, I have justified their commitment and maybe even earned their respect.

This journey would not have been possible without the unwavering love and support of my wife, Kim. A part-time PhD is a marathon for the candidate, but one that is self-imposed. The burden is different for a partner, no less demanding, and often overlooked. While I was consumed by research and writing, Kim carried the heavier load at home, raising our sons, holding our lives together, and bearing the emotional weight of my absences, both physical and mental. Her patience, humour, strength, and faith in me never faltered. This thesis bears my name, but it belongs to her as much as it does to me.

Completing this thesis has been intellectually demanding, personally revealing, humbling, and rewarding in ways I could not have anticipated. It marks the end of one journey and the quiet beginning of another. I finish with no illusions, only gratitude for the knowledge, lessons, and people who carried me through.

Professional editor Josephine Smith – WordSmith WA provided copyediting and proofreading services in accordance with the guidelines laid out in the university-endorsed *Guidelines for Editing Research Theses* (revised 25 February 2019).

Table of Contents

List of Figures

List of Tables

Acronyms and abbreviations

ABM: Agent-based modelling

ACARA: Australian Curriculum, Assessment and Reporting Authority

ANN: Artificial Neural Network

AUC: Area Under the Curve

BIC: Bayesian Information Criterion

CESE: Centre for Education Statistics and Evaluation

CSV: Comma-Separated Values

EBPM: Evidence-Based Policymaking

EDA: Exploratory data analysis

GERM: Global Education Reform Movement

GBT: Gradient boosted tree

GDP: Gross domestic product

GUI: Graphical User Interface

ICSEA: Index of Community Socio-Educational Advantage

ILSA: International large-scale assessment

LIME: Local Interpretable Model-Agnostic Explanations

NAPLAN: National Assessment Program – Literacy and Numeracy

NPM: New Public Management

NSIP: National School Improvement Partnerships

NSW: New South Wales

OECD: Organisation for Economic Co-operation and Development

PCA: Principal Component Analysis

PPIPA: Privacy and Personal Information Protection Act 1998 (NSW)

PISA: Programme for International Student Assessment

RCT: Randomised control trial

ROC: Receiver Operating Characteristic

SDM: Systems Dynamics Modelling

SEF: School Excellence Framework

SHAP: SHapley Additive exPlanations

SINSW: School Infrastructure New South Wales

SOC: School Organisational Climate

SSM: School Success Model

SVM: Support vector machine

t-SNE: t-Distributed Stochastic Neighbour Embedding

TIMSS: Trends in International Mathematics and Science Study

WHITS: What's Happening In This School

Glossary of terms

Causal loop diagrams: A tool used in systems dynamics to represent feedback loops and relationships in complex systems.

Cluster centroid: The central vector or average of a cluster in multi-dimensional space, representing the cluster's central characteristics.

Clustering for utility: A methodological approach that uses clustering to create abstract, nominal class variables for enabling predictive analysis.

Complex and dynamic systems: Systems characterised by interconnected elements and unpredictable behaviour, such as schools.

Evidence-Based Policymaking (EBPM): An approach focused on reductionist analysis and assumptions of objectivity, replicability, and causality. It prioritises generalisability and predictability, in contrast to adaptive or systems-based methodologies that emphasise emergent, relational, and context-specific factors in complex environments.

Gradient boosted trees (GBTs): A machine learning technique for classification and regression that builds predictive models from sequentially assembled decision trees.

High-stakes assessment: Evaluations of school-level performance with significant implications for students, educators, and institutions, influencing funding, progression, accountability, and policy decisions.

Index of Community Socio-Educational Advantage (ICSEA): A metric that indicates the socio-educational advantage of a school compared to others, based on parental education, occupation, and school location.

Institutional Environment: The physical and sensory aspects of a school, including infrastructure, maintenance, and ergonomics.

International large-scale assessments (ILSAs): Internationally comparable assessments of academic performance.

K-means clustering: A partitioning algorithm that divides data into a predefined number of clusters by iteratively assigning data points to the nearest centroid and then updating the centroids to minimise intra-cluster variance.

Predictive weights: The importance assigned to variables in predictive modelling, reflecting their contribution to prediction accuracy.

Receiver Operating Characteristic (ROC) curve: A graphical plot illustrating the diagnostic ability of a binary classifier system.

School Climate: The quality and character of school life, encompassing the four pillars: Safety, Interpersonal Relationships, Teaching and Learning, and Institutional Environment.

School Climate misfit: The difference between the students' current and preferred school experiences, measured on a Likert scale.

School effect: Variations in educational outcomes attributable to differences in school environments rather than individual or family characteristics.

System dynamics models: Models used to simulate and analyse the behaviour of complex systems over time.

X-means clustering: An extension of k-means clustering that determines the optimal number of clusters based on the data.

Chapter 1. Introduction

Educational systems worldwide face increasing pressure to improve student outcomes, driven by heightened global competition (Lingard & Rizvi, 2009) and the widespread and growing use of standardised assessments (Minarechová, 2012). These pressures have increasingly led researchers and policymakers to adopt policymaking approaches that identify and leverage measurable relationships between educational inputs, such as teacher quality, curriculum design, resource allocation, and outputs, including assessed students' literacy and numeracy levels. These approaches are categorised as Evidence-Based Policymaking (EBPM) (Geyer, 2012). However, while this approach continues to influence educational policy, it often prioritises easily quantifiable factors, overlooking less direct but potentially critical elements. Further, the effectiveness of EBPM in achieving demonstratable, system-wide improvements is increasingly questioned (Biesta, 2010).

This PhD research highlights the potential for techniques used in EBPM to overlook or underestimate the role of school infrastructure in shaping educational outcomes (Crampton, 2009). Researchers have identified that infrastructure can impact outcomes through indirect and interconnected relationships (Uline & Tschannen-Moran, 2008; Buckley, et al., 2005). Many analytical approaches used in EBPM prioritise direct cause-and-effect relationships (Bromley, et al., 2023), potentially undervaluing more complex interactions between inputs and outputs (Bloom, 2019; Bergeron & Rivard, 2017).

School infrastructure, encompassing buildings and the broader built environment, is a prominent feature of educational systems and requires substantial initial and ongoing financial investment (Hong & Zimmer, 2016). Policymakers typically regard infrastructure as necessary for enabling education but as peripheral to improving educational outcomes (Gunter & Shao, 2016). As such, despite being central to the educational experience and a potential influence on outcomes (Wang & Degol, 2016), policymakers often overlook infrastructure as a lever for improving educational outcomes (Belmonte, et al., 2019). This view of infrastructure reflects a dominant belief within the evidence-based paradigm that infrastructure exerts only a minor influence on student performance (Martorell, et al., 2016; Crampton, 2009). This view persists even as emerging studies suggest that infrastructure impacts key aspects of school experience, including student wellbeing (Maxwell, 2016), teacher satisfaction (Buckley, et al., 2005), and academic performance (Uline & Tschannen-Moran, 2008). However, these studies are often fragmented (Crampton, 2009) and little research has systematically investigated the predictive power of infrastructure in shaping student outcomes when considered within a broader framework of school experience such as School Climate.

In parallel with the growing dominance of EBPM (Biesta, 2010; Brown, 2018), educational policy has increasingly relied on and been influenced by performance benchmarking against standardised assessments of student academic achievement (Bromley, et al., 2023). Since its introduction in 2000, the *Programme for International Student Assessment* (PISA) has emerged as the dominant measure of cross-national educational performance (Breakspear, 2012), with the published results influencing educational policy across the 38 Organisation for Economic Co-operation and Development (OECD) member states and beyond (World Population Review, 2024). In addition to PISA, assessments such as the *Trends in International Mathematics and Science Study* (TIMSS) and, specific to Australia, the *National Assessment Program – Literacy and Numeracy* (NAPLAN) have further entrenched policymakers' focus on quantifiable academic performance. The widespread reliance on these high-stake assessments reflects the prevailing assumption that academic achievement is a proxy for future national economic competitiveness (Chen & Luoh, 2009), individual lifetime earnings (Balestra & Backes-Gellner, 2017), and even longevity (Heckman, et al., 2018). Consequently, policymakers have prioritised interventions intended to enhance students' literacy and numeracy skills, often at the expense of more holistic educational considerations such as infrastructure, wellbeing, and school culture (Morsy, et al., 2018; OECD, 2018).

Despite the significant cost to establish and operate school infrastructure, it remains marginalised in policy debates regarding school improvement. New South Wales (NSW) has approximately 2,200 schools, housing 27,500 buildings and 45,000 classrooms. The 2024/2025 NSW budget allocated AUD$8.9 billion to construct new schools and upgrades, representing roughly one-third of the state's total education budget (NSW Department of Education, 2024). However, the Centre for Education Statistics and Evaluation (CESE), the research body within the NSW Department of Education, does not consider infrastructure a key determinant of school performance, concluding that there is "little or mixed" evidence supporting its impact beyond basic environmental factors such as lighting, noise, and ventilation (2015, p. 15). Further, infrastructure is absent in NSW Department of Education performance improvement doctrine (NSW Department of Education, 2017; NSW Department of Education, 2020).

This study situates infrastructure within the School Climate construct. School Climate is a framework used to describe and categorise school experience (Thapa, et al., 2013). Researchers have found that improvements in School Climate predict improvements in student outcomes (Hopson & Lee, 2011; Reynolds, et al., 2017). Increasingly, researchers consider infrastructure as important to School Climate, whether directly as part of the construct (Wang & Degol, 2016) or indirectly (Uline & Tschannen-Moran, 2008).

1.1 Research problem

EBPM, increasingly dominant in educational and other public policy development, aligns with a traditional scientific worldview, emphasising causality, reductionism, predictability, and determinism (Geyer, 2012). Regression analysis, commonly employed in EBPM, is well suited to quantifying simple, linear cause-and-effect relationships between isolated variables (Wu & Coggeshall, 2012). However, the dominant techniques underpinning EBPM are not well suited to capturing the interconnected and evolving nature of educational environments (Farhood, et al., 2024), which exhibit characteristics of complex and dynamic systems (Goldspink, 2007; Mason, 2016).

Schools are not mechanistic systems where standardised inputs yield predictable outputs (Koopmans, 2020); instead, they are adaptive, evolving environments (Rudasill, et al., 2018) where interactions between students, teachers, infrastructure, and policies generate emergent and often unpredictable outcomes (Bloom, 2019). However, policymakers continue to rely on regression-based analyses, which overlook the non-linear and interdependent nature of school dynamics (Fullan, 2019), and therefore the resulting educational interventions typically fail to achieve anticipated improvements in student performance (Biesta, 2010; Waslander, et al., 2016). One consequence of this methodological approach is the underexplored role of school infrastructure in influencing student experiences and outcomes (Crampton, 2009). Infrastructure, typically regarded as a passive background factor, has generally been marginalised in policy discussions regarding school improvement (Holmund, et al., 2010). The prevailing assumption is that infrastructure contributes only marginally to educational outcomes, exerting a limited and mainly indirect influence on student performance (Hattie, 2023; Martorell, et al., 2016). This assumption persists despite growing evidence that infrastructure shapes key aspects of student experience, including safety, comfort, engagement, and school climate (Uline, et al., 2010). Much of the existing research focuses on infrastructure as an independent variable, assessing its impact in isolation rather than considering how it interacts with broader educational factors (Gunter & Shao, 2016). This approach fails to consider how infrastructure, as an element of the broader school environment, may be a key mediator of student experience and learning outcomes (Buckley, et al., 2005).

Instead of viewing infrastructure as a static input to education, this research considers it to be a component of a complex and dynamic system (Steen, et al., 2013), where its effects emerge through interdependencies with other elements of the school environment (Uline & Tschannen-Moran, 2008). Accordingly, this study seeks to understand the contribution of school infrastructure to educational outcomes within the context of complexity theory, which emphasises the interconnected and evolving nature of educational environments (Koopmans, 2020).

Rather than treating infrastructure as a discrete variable with a fixed and isolatable contribution to student achievement, this study explores how students' lived experience of their physical learning environment interacts with the broader dimensions of School Climate, such as perceptions of safety, quality of interpersonal relationships, and teaching practices (Thapa, et al., 2013). This reframing reflects a theoretical shift from linear causality (Ansell & Geyer, 2017) towards an appreciation of the emergent, systemic nature of educational outcomes (Rudasill, et al., 2018; Slavin, 2008).

Although the School Climate framework inherently acknowledges the interconnectedness of school factors (Wang & Degol, 2016), much of the empirical research remains grounded in regression-based methodologies that privilege independent, additive effects (Urick & Bowers, 2014; Aldridge, et al., 2024). The continued reliance on statistical techniques such as multiple regression, hierarchical linear modelling, and other variable-centred methods that deconstruct complex relational phenomena into discrete, linear associations (Wrigley, 2019) may therefore be misaligned with the holistic and dynamic character of School Climate. This methodological simplification, generally referred to as a reductionist orientation (Chen, et al., 2024), risks fragmenting inherently interconnected school processes into isolated variables, thereby constraining researchers' ability to capture, interpret, and theorise the full complexity of educational environments (Bloom, 2019; Zadza, 2020).

Informed by complexity theory (Mitchell, 2009), this study challenges the assumption that school improvement can be meaningfully driven by isolating discrete variables and identifying singular causal pathways (Biesta, 2010). This study considers how infrastructure contributes to the overall climate of a school (Wang & Degol, 2016) and how complex interactions shape student experiences and outcomes (Steen, et al., 2013). In doing so, this research offers a theoretically grounded alternative to reductionist models of school improvement (Geyer, 2012; Bloom, 2019) by advocating for a more holistic, contextually sensitive approach to understanding the role of infrastructure in educational settings.

1.2 Objectives and the research question

This study explores the complex relationship between the students' experiences of school infrastructure and educational outcomes within the dynamic ecosystems of contemporary educational environments. Despite significant investment in educational infrastructure, its impact on student outcomes remains underrepresented, mainly in policy discussions and academic research. Researchers often treat infrastructure as a discrete input, primarily assessed through expenditure data or condition surveys and analysed through linear statistical models that assume stable, direct relationships between inputs and outputs (Crampton, 2009; Hong & Zimmer, 2016; Geyer, 2012). These assumptions can obscure the interdependent and context-sensitive nature of school systems.

Therefore, the primary research question guiding this investigation is:

RQ: *What is the relationship between school infrastructure and educational outcomes?*

To deepen the inquiry, this study poses two subsidiary questions. The first subsidiary question is:

SRQ 1: *How does situating infrastructure within the broader context of student experience inform the understanding of its role in shaping educational outcomes?*

This question reconceptualises infrastructure not merely as a static or technical input but as a critical element of the students' lived experiences within the school environment, interacting meaningfully with other experiential elements such as Safety, Interpersonal Relationships, and Teaching and Learning. This framing uses the School Climate model, which conceptualises student experience as multidimensional and interdependent (Wang & Degol, 2016).

The second subsidiary question is:

SRQ 2: *To what extent does a complexity-informed approach help to assess the contribution of infrastructure to educational outcomes?*

This question reflects the study's analytical orientation. Grounded in complexity theory, it positions schools as adaptive systems in which outcomes emerge from the intricate interactions of multiple, evolving components (Mitchell, 2009; Mason, 2016). This perspective facilitates non-linear, relational interpretations of the influence of infrastructure, moving beyond the constraints of traditional input–output models.

Together, these questions establish a comprehensive research agenda. They emphasise the study's intention to move beyond reductionist paradigms; use analytical tools that align with the complexities of systemic interactions; and cultivate a more integrated understanding of infrastructure within the ecology of school life. By addressing these questions, this research aims to enrich the discourse on educational outcomes and infrastructure, ultimately contributing to more effective policy and practice in educational settings.

1.3 Significance of the study

This study makes significant theoretical, methodological, and practical contributions to understanding the contribution of infrastructure to educational outcomes within the context of complex and dynamic systems.

The first theoretical contribution of this study the operationalisation of infrastructure within the Institutional Environment pillar of School Climate model, first proposed by Wang and Degol (2016). This provides the initial foundation for measuring the effect of infrastructure as part of the broader School Climate construct. Crampton (2009) noted that changes in infrastructure affect the efficacy of other changes, which is consistent with the view of Gunter and Shao (2016) that the direct measured contribution of infrastructure may not reflect the actual contribution. A noted attribute of many studies examining the effect of infrastructure is the focus on expenditure as a proxy for improvement (Belmonte, et al., 2019; Hong & Zimmer, 2016), potentially compromising efforts to compare infrastructure and other changes (Maxwell, 2016). Operationalising infrastructure in the School Climate framework using a single survey instrument moves to address the noted shortcomings in understanding the effect of infrastructure on educational outcomes (Crampton, 2009).

The second theoretical contribution is applying cluster analysis to define educational outcomes using student-reported wellbeing and behaviour. Wellbeing and behaviour are strongly associated with long-term educational attainment, which is a significant predictor of longevity (Ross & Wu, 1995; Hart, et al., 2017), quality of life (Gutacker, et al., 2023), and self-determination (Bergman & Hupka-Brunner, 2013). Specifically, this study uses cluster analysis to group students according to the overall outcomes reported. The application of clustering analysis in the context of student outcomes is not unique, with numerous studies using clustering to identify outcome groups (Dumuid, et al., 2017; Gonzalvez, et al., 2018), which are later dissected and analysed. Where this study is unique is in the use of the outcome clusters as the target for later prediction without the need to isolate the individual contributing factors, which is a noted weakness in School Climate analysis (Wang & Degol, 2016).

The third theoretical contribution of this study is to apply complexity theory directly to the conceptualisation and analysis of School Climate. Although School Climate frameworks aim to provide a broad understanding of schools by accounting for diverse experiences and interactions (Rudasill, et al., 2018), empirical studies in this area have generally remained reliant on traditional, linear methodologies. These reductive approaches, which treat school factors as independent and additive variables, risk undermining the relational and dynamic nature that the School Climate framework seeks to capture. At the same time, complexity theorists have often rejected categorisation frameworks such as School Climate (Van Houtte & Van Maele, 2011; Feldhoff, et al., 2021), regarding them as not only overly rigid but also too narrow in scope to account for the emergent and adaptive properties of real-world systems (Rudasill, et al., 2018). This study, however, argues that School Climate and complexity theory are compatible. Bringing complexity theory to bear on School Climate offers an alternative to linear models, which are dominant in educational research and EBPM (Geyer, et al., 2015). Recognising schools as complex, adaptive systems, this study recognises the dynamic and reciprocal interactions between infrastructure, student experience, and broader School Climate factors (Maxwell, 2016). This repositioning of School Climate in the context of complexity challenges reductionist policy approaches that conceptualise infrastructure as a discrete and static variable contributing independently to educational outcomes (Bromley, et al., 2023; Crampton, 2009), but instead frames infrastructure as an integral, interdependent element embedded within school experience (Wang & Degol, 2016). Integrating complexity theory into the study of School Climate thus provides a richer, more dynamic framework for understanding educational environments, offering both theoretical and practical advantages.

Methodologically, this study contrasts with traditional approaches used to influence school improvement policy (Biesta, 2010) by employing advanced data analysis techniques, including machine learning, to investigate non-linear relationships between dependent and independent variables. This approach responds to a growing recognition in the literature that schools operate as complex, dynamic systems (Koopmans, 2020) with school-specific patterns (Sellström & Bremberg, 2006), contextual influences, and reliance on interactions between school users and their environment (Rudasill, et al., 2018; Steen, et al., 2013).

Unlike traditional statistical techniques such as regression analysis, which assume stable, linear relationships, machine learning allows for the identification of emergent patterns, interactions, and feedback loops that better reflect the realities of school environments (Zaman, et al., 2023). By applying machine learning algorithms, this study advances the development of predictive models that are sensitive to the complex, dynamic, and school-specific patterns characteristic of real-world educational environments (Zaman, et al., 2023).

Practically, the findings may inform more informed and tailored policy considerations and decisions regarding resource allocation and school improvement strategies. A nuanced understanding of the role of infrastructure in educational outcomes can provide evidence for targeted investments that support a broader, contextually appropriate and responsive approach to improving School Climate and, therefore, student wellbeing and behaviour.

In summary, this study advances a new approach to evaluating school infrastructure, recognising its indirect influence on student outcomes within complex, dynamic systems. It critiques the limitations of traditional evidence-based methodologies and reconceptualises infrastructure as an active element of school experience. By integrating complexity theory, machine learning, and the School Climate framework, it challenges prevailing assumptions in educational policy and lays the groundwork for future research centred on schools' emergent, systemic properties.

1.4 Structure of the thesis

This introductory chapter articulates the research problem, establishes the objectives, and outlines the key research question guiding the study. By emphasising the significance of the research, this chapter sets the stage for a rigorous exploration of the interplay between educational assessments, policymaking, and school infrastructure. The subsequent chapters of this thesis are structured to provide a comprehensive understanding of these dynamics. The structure of the thesis is shown in Figure 1.1.

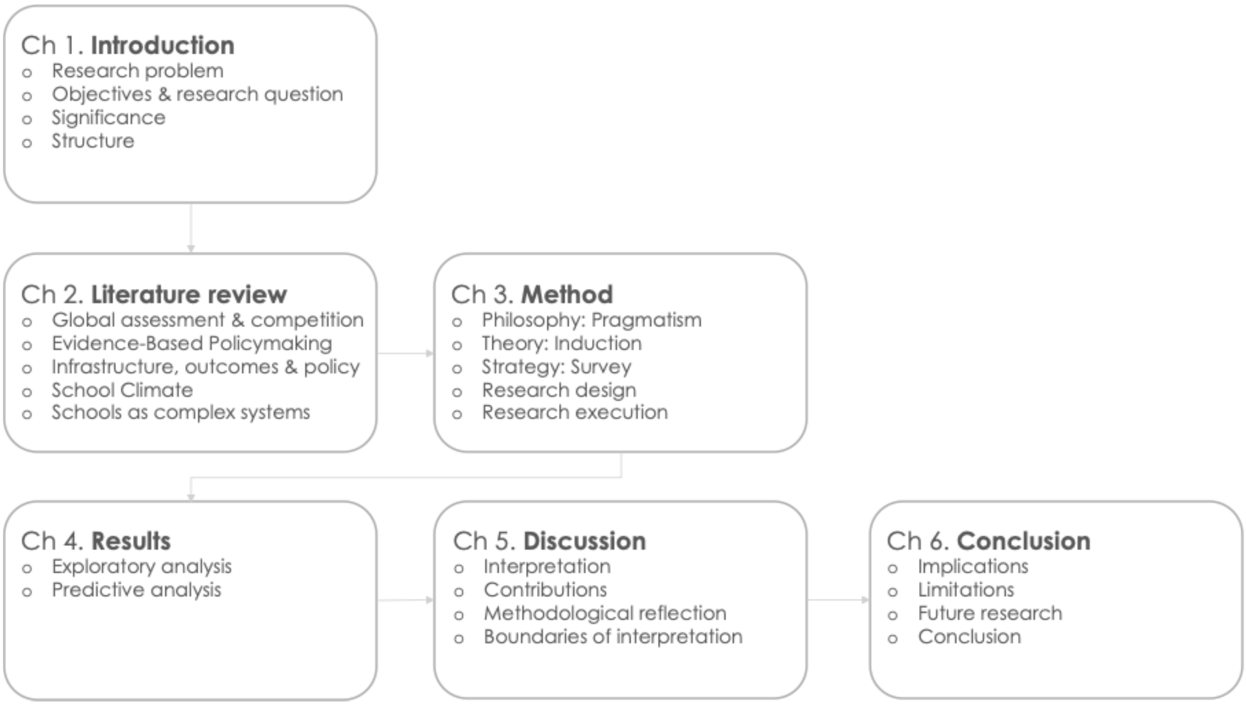

Figure 1.1. Thesis structure

Following the Introduction, Chapter 2 reviews the existing literature by examining critical themes including global educational assessments, EBPM, and the relationship between School Climate and educational outcomes. Following this, Chapter 3 presents the methodological framework underpinning the research and details the philosophical stance of pragmatism; the adoption of inductive theory-building; and a clear outline of the research design and survey strategy. Chapter 4 presents the study's findings, beginning with an exploratory analysis of the data before transitioning to the outcomes of cluster analysis and predictive modelling. Chapter 5 critically discusses and reflects on the research findings and explores their interpretations and contributions to the field while contemplating the methodological considerations and the scope of interpretation. Finally, Chapter 6 concludes the thesis by examining the broader implications of the research findings, acknowledging the limitations, and proposing avenues for future inquiry.

Chapter 2. Literature review

This chapter reviews the research on the role of school infrastructure in shaping educational outcomes, situating this work within broader discussions of School Climate, educational policy, and systemic influences on learning. The literature review is organised around five thematic explorations:

1. The influence of standardised assessments and EBPM.
2. The overlooking of infrastructure in policymaking.
3. The insights from School Climate research.
4. The limitations of traditional analytical methods in informing educational policy.
5. The relevance of complexity theory.

The review highlights the potential for a nuanced and dynamic understanding of how infrastructure contributes to educational outcomes within complex school environments.

The review situates this inquiry within the rise of standardised assessments (Chezan, et al., 2017) and the emergence of EBPM (Ansell & Geyer, 2017; Zadza, 2020), which have profoundly influenced how educational success is defined and pursued globally (Bromley, et al., 2023). A central concern is the dominance of EBPM, which typically relies on linear causal assumptions linking inputs, such as teacher quality, curriculum, and funding, to outputs, which are typically measured through literacy and numeracy scores (Cartwright & Hardie, 2012). Although such relationships are frequently statistically significant (Wiseman & Davidson, 2018), their practical impact on improving educational outcomes remains contested (Minarechová, 2012; O'Neill, 2012). Within this paradigm, the contribution of infrastructure is often marginalised (Crampton, 2009) despite substantial global investment (Gunter & Shao, 2016).

School Climate research offers an alternative perspective on the role of infrastructure as a contributor to educational outcomes (Wang & Degol, 2016), emphasising the importance of student experience and highlighting how perceptions of the built environment influence outcomes (Uline & Tschannen-Moran, 2008; Thapa, et al., 2013). Studies outside the realm of School Climate have also demonstrated that infrastructure can affect student behaviour (Berman, et al., 2018) and teacher satisfaction (Roberts, 2009), two key predictors of academic success (Anderson, 2005).

While research supporting EBPM and School Climate conceptualised schools differently, both predominantly rely on linear, regression-based analyses (Snyder, 2013), limiting their capacity to account for the non-linear and emergent dynamics characteristic of educational settings (Koopmans, 2020). A growing body of research reflects a different approach where schools are considered as complex, adaptive systems (Goldspink, 2007; Waslander, et al., 2016), suggesting that systemic properties may explain the persistent challenges in EBPM based educational reform (Biesta, 2010; Geyer, 2012).

This chapter critically examines the prevailing methodologies and theoretical frameworks underpinning research on school infrastructure and educational outcomes. This chapter concludes by synthesising the existing research, examining the relevant gaps, and proposing the research question.

2.1 Global assessment, comparison, and competition driving educational reform

The expansion of ILSA regimes has transformed global educational policy (Sahlberg, 2007; Sayer, 2020) and promoted a data-driven approach to system improvement (Brown, 2018). Nations increasingly prioritise measurable outcomes to enhance their rankings, leading to widespread reforms focused on performance metrics rather than broader educational aims (Bromley, et al., 2023). This trend has positioned large-scale policy interventions as the primary means of educational improvement, which often filters down to subnational levels such as states, provinces, and local authorities (Wiseman & Davidson, 2018).

International large-scale assessments (ILSAs) such as PISA have become central to evaluating and reforming educational systems (Lingard, et al., 2013; Breakspear, 2012). Initially designed as comparative benchmarking tools (OECD, 2020), these assessments have evolved into influential policy instruments that shape national curricula, accountability frameworks, and public discourse regarding educational performance (Sellar & Lingard, 2014; Reid, 2020; Putansu, 2020). Governments now view participation in ILSAs as crucial for demonstrating system effectiveness in a competitive global context (Lingard & Rizvi, 2009). Consequently, test rankings frequently dictate perceptions of educational policy success or failure (Minarechová, 2012; Bonal & Tarabini, 2013) and drive reform efforts to improve measurable outcomes rather than address deeper educational needs (Breakspear, 2012; Morgan, 2014).

A principal motivation for the dominance of ILSAs is the belief that test performance correlates with national economic success (Baldwin, et al., 2011). Policymakers assert that higher rankings indicate future workforce productivity and global competitiveness (Bonal & Tarabini, 2013), reinforcing the idea that educational systems should prioritise economic growth (Brunello & Paola, 2014). However, the empirical evidence for this assertion is contested, with Chen and Luoh (2009) finding no consistent link between PISA scores and long-term gross domestic product (GDP) growth. This finding is consistent with Zadza, (2020), indicating that the relationship between standardised test results and economic outcomes may be oversimplified. Moreover, ILSAs provide governments with politically advantageous metrics for justifying policy decisions (Lingard & Rizvi, 2009; Ball, 2012), with test rankings offering clear, quantifiable indicators that portray educational performance as either a crisis or a success (Wiseman & Davidson, 2018). This dynamic renders educational policy increasingly reactive to ranking fluctuations (Geyer, 2012), prompting governments to implement large-scale reforms in response to perceived underperformance, often without critically evaluating the efficacy of these interventions (Morgan, 2014; Putansu, 2020).

As ILSA rankings gain prominence, policy borrowing and standardisation have emerged as pervasive trends in education governance. Nations increasingly emulate reforms from high-performing systems, such as Finland, Singapore, and South Korea, but frequently overlook contextual differences (Sahlberg, 2021). The dominance of ILSAs has fundamentally reshaped the evaluation and governance of educational systems (Minarechová, 2012). A global reform movement has emerged that prioritises performance metrics over broader educational objectives and positions large-scale policy interventions as primary tools for enhancing system effectiveness (Bromley, et al., 2023). Crucially, this transformation stems from the rise of EBPM, which endorses data-driven decision-making (Morrison, 2021), standardised indicators, and measurable impacts (McKnight & Morgan, 2020) as the foundation of educational reform (Lobascher, 2011). The following section explores EBPM in more detail, elucidating its core assumptions, methodologies, and influence on policy and practice in education.

2.2 The rise and dominance of Evidence-Based Policymaking in educational reform

The increasing emphasis on measurable educational outcomes has positioned EBPM as a central mechanism in global educational reform (Geyer, 2012). Governments and policymakers have relied on empirical research, large-scale data analysis, and systematic evaluation to guide policy decisions (Chezan, et al., 2017; Lingard, 2013), with the expectation that this approach improves educational outcomes (O'Neill, 2012). The expansion of ILSA regimes such as PISA and TIMSS has reinforced this trend by providing quantifiable benchmarks against which national educational systems can be compared and evaluated (Breakspear, 2012; OECD, 2020). As a result, EBPM has not only enabled but also actively accelerated global educational reform (Wiseman & Davidson, 2018; Gorard, 2020).

The emergence of EBPM as a dominant policy framework is rooted in broader governance trends, particularly the rise of New Public Management (NPM) in the late 20th century (Christensen & Laegreid, 2022). The shift towards data-driven decision-making, accountability, and performance measurement reflected a growing belief that policy should be guided by objective, research-based evidence rather than political ideology or professional intuition (Biesta, 2010; Geyer, et al., 2015; Gorard, 2020). Educational policymakers, influenced by this movement, increasingly sought to identify “what works best” (Fuller, 2022, p. 99) through quantitative research, meta-analysis, and large-scale empirical studies (Fullan, 2019). However, while EBPM has provided a systematic framework for policy formulation, it has also been criticised for its over-reliance on quantifiable data; its methodological limitations (Brown, 2018; Bergeron & Rivard, 2017); and its frequent disregard for contextual complexities

(Cartwright & Hardie, 2012). Despite these critiques, EBPM proponents argue that its reliance on rigorous empirical evidence enhances transparency and accountability, which ensures that policies are based on proven effectiveness rather than anecdotal experience or political motivations (Lingard, et al., 2013).

This section explores the rise and dominance of EBPM in educational policy (Zadza, 2020), outlining its theoretical foundations (Sayer, 2020), the methodological tools that sustain it (Simpson, 2018), and its practical implications in educational reform (Putansu, 2020). The analysis examines the core statistical techniques used in EBPM before critically examining the dominant statistical techniques and the role of meta-analysis (Simpson, 2017) as a key mechanism within EBPM. This review then transitions to an analysis of the role of infrastructure within the EBPM approach and examines how the dominant analytical methodologies have shaped perceptions of infrastructure investment (Crampton, 2009; Gunter & Shao, 2016) and its perceived impact on educational outcomes. This section then explores how EBPM has shaped policy implementation globally (Bromley, et al., 2023), specifically focusing on New South Wales. Finally, the discussion considers the inherent challenges of applying EBPM in complex environments (Ansell & Geyer, 2017), including education (Mason, 2016), setting the stage for a broader critique of its universal applicability.

2.2.1 The foundations of Evidence-Based Policymaking

EBPM is underpinned by a scientific worldview, which assumes that educational phenomena can be objectively measured, quantified, and analysed using scientific methods (Geyer, 2012). This approach aligns with positivist epistemology, which posits that reality is orderly, predictable, and governed by causal relationships that can be discovered through systematic observation and experimentation (Sayer, 2020). Advocates of EBPM argue that rigorous statistical techniques and experimental methodologies can identify the most effective policy interventions (Putansu, 2020; Hattie, 2013). This belief has led to the privileging of quantitative research methodologies such as randomised controlled trials (RCTs) (Castillo & Wagner, 2014), regression-based causal inference, and meta-analysis (Morrison, 2021) as these approaches provide numerical representations of policy effectiveness, which are presumed to be more reliable and generalisable than qualitative insights (Cartwright & Hardie, 2012). The influence of this perspective is evident in global educational policymaking (Bromley, et al., 2023; Reid, 2020), where quantitative indicators such as standardised test scores and graduation rates are frequently used to evaluate system performance and guide reform initiatives (Breakspear, 2012; Minarechová, 2012; Morsy, et al., 2018).

However, critics argue that this scientific worldview overlooks education’s social, cultural, and contextual dimensions (Gorard, 2013; Geyer, 2012), which are difficult to quantify and may not conform to the linear cause-and-effect relationships assumed in EBPM (Blass, 2020). The positivist assumption that objective truths about education can be uncovered through quantitative analysis (Waslander, et al., 2016) fails to take into account the complexity of learning environments (Fullan, 2019); the role of teacher–student interactions (Rudasill, et al., 2018); and the impact of socioeconomic factors on student achievement (Lingard, 2013). Furthermore, the reliance on large-scale datasets and statistical modelling may obscure the lived experiences of students and teachers, reducing education to an abstract system of inputs and outputs rather than a dynamic, human-centred process (Gorard, 2013).

Despite these critiques, EBPM remains a powerful force in global education governance (Ansell & Geyer, 2017), shaping funding allocations, curriculum design, and teacher evaluation systems (Lingard, et al., 2013). The dominance of this paradigm has been reinforced by international organisations such as the OECD (OECD, 2020) and the World Bank (World Bank Group, 2025), which promote evidence-based strategies as the gold standard for educational policymaking (Sellar & Lingard, 2014). While the scientific approach underpinning EBPM has contributed to greater accountability and transparency in educational policy (Wiseman & Davidson, 2018), its limitations highlight the value of a nuanced approach to understanding educational effectiveness that incorporates qualitative perspectives, practitioner insights, and contextual considerations (Biesta, 2010; Head, 2016; Fullan, 2019)

2.2.2 The techniques underpinning Evidence-Based Policymaking

EBPM relies on a range of statistical techniques to determine the effectiveness of educational interventions. These techniques establish causal relationships, measure intervention impact, and control for confounding variables (O'Neill, 2012). One of the most widely used statistical approaches in EBPM is regression analysis (Geyer, 2012), which enables researchers to estimate the relationship between multiple educational variables while controlling for extraneous influences (Benkhalfallah & Laouar, 2023). Multiple regression models are frequently employed to predict student performance based on socioeconomic status, school funding, teacher quality, and curriculum interventions (Parker, et al., 2018). However, critics argue that the assumption of linear relationships between variables oversimplifies the complexity of educational systems (Zengaro & Warley, 2022) as student learning is influenced by interdependent, non-linear factors that may not be adequately captured in standard regression models (Waslander, et al., 2016).

A statistical method commonly used in EBPM is RCT (Castillo & Wagner, 2014), which is considered the gold standard for causal inference (McKnight & Morgan, 2020). This method has been widely used to evaluate literacy programs, teacher training initiatives, and classroom interventions (Morrison, 2021). Despite their methodological rigour, RCTs face criticism when applied to improving educational outcomes, with researchers noting that controlled experimental settings often fail to replicate the dynamic, multifaceted nature of real-world school environments (Cartwright & Hardie, 2012). Nonetheless, EBPM proponents emphasise that RCTs remain one of the few methods capable of demonstrating causal relationships, which makes them indispensable for guiding large-scale policy initiatives (Gorard, 2020).

Effect size calculations are standard in EBPM, allowing policymakers to compare the relative impact of different interventions across multiple studies (Simpson, 2018). Effect size measures, such as Cohen's and Hedges' (Taylor & Alanzi, 2023), provide a standardised metric for assessing the magnitude of an educational intervention's effect (Marfo & Okyere, 2019). Effect size calculations are useful in meta-analysis, enabling researchers to synthesise findings from diverse studies and generate overall conclusions regarding intervention efficacy (Velasco-Benítez & García-Perdomo, 2019). However, critics argue that effect size calculations can be misleading if the underlying studies differ in quality, sample size, or research design, which make direct comparisons problematic (Simpson, 2018).

2.2.3 The contribution of meta-analysis to Evidence-Based Policymaking

Meta-analysis is a statistical method used to integrate results from multiple studies, enabling researchers to draw generalised conclusions about effective practices in education (Zadza, 2020). The rationale behind this method is that by aggregating and analysing large bodies of research, policymakers can make more informed decisions based on the weight of empirical evidence rather than isolated studies (Bergeron & Rivard, 2017).

Meta-analysis has significantly influenced educational research and policymaking because it offers a systematic and quantifiable approach to evaluating interventions (Polesel, et al., 2014). The assumption underpinning its use is that by distilling complex educational data into effect sizes (Simpson, 2018) and rankings, policymakers can identify the most effective strategies for improving student learning outcomes (Sayer, 2020). This assumption has led to the widespread adoption of meta-analytical findings in teacher evaluation, funding decisions, and large-scale educational reforms (Blanchenay, et al., 2014).

A landmark meta-analysis study, both in scale and influence, is Hattie's Visible Learning (2023), which synthesised over 800 meta-analyses, encompassing more than 50,000 studies and 300 million students worldwide. Hattie ranked various educational interventions by effect size, establishing a "hinge point" of 0.4 as the threshold for educational effectiveness (Hattie, 2023). Hattie's work has been widely adopted in teacher training, curriculum development, and school improvement models in Australia (Blass, 2020), New Zealand (O'Neill, 2012), and Denmark (Knudsen, 2017).

Despite its prevalence and influence, meta-analysis has also been subject to substantial critique. Scholars have raised concerns about its tendency to oversimplify educational phenomena by reducing diverse and context-dependent interventions into a single numerical effect size (Simpson, 2018; O'Neill, 2012). Critics argue that combining studies with different methodologies, populations, and research designs can lead to misleading conclusions if suitable methodological controls are not applied (Bergeron & Rivard, 2017; Blass, 2020). Moreover, publication bias, where studies with significant findings are more likely to be published and included in meta-analyses, has led to overestimating the effectiveness of specific interventions (Lin & Chu, 2018).

2.2.4 Theoretical criticisms of Evidence-Based Policymaking

Despite the extensive use of statistical techniques in EBPM, scholars have identified several key limitations in their application. One of the most frequently cited issues is the tendency to assume that correlation implies causation, which can lead policymakers to implement interventions based on statistical associations rather than demonstrated causal mechanisms (Cartwright & Hardie, 2012). This practice is particularly problematic in observational studies, where unmeasured confounding variables may influence results, leading to spurious policy recommendations (Foster, 2023).

Another major limitation is the over-reliance on quantitative indicators at the expense of qualitative insights (Biesta, 2010). While statistical models and meta-analyses provide numerical assessments of policy effectiveness (Parker, et al., 2018), they often fail to capture the social, cultural, and pedagogical nuances of educational settings (Koopmans & Stamovlasis, 2016). As a result, policies based solely on quantitative findings may overlook important contextual factors such as teacher expertise, school leadership, and community engagement (Morsy, et al., 2018), which play a crucial role in educational success (Loughland & Thompson, 2016). Alternative approaches such as qualitative research and mixed-methods studies may provide deeper insights into these complexities; however, they remain underused in EBPM frameworks (O'Neill, 2012; Wrigley, 2019).

Despite these criticisms, proponents of EBPM maintain that statistical methods provide the most reliable foundation for policymaking, ensuring that educational reforms are guided by empirical evidence rather than anecdote or ideology (Putansu, 2020).

2.2.5 Criticisms of Evidence-Based Policymaking in application

Despite its widespread adoption, EBPM in education has been subject to extensive critique, particularly regarding its practical application in diverse educational contexts. Scholars argue that while EBPM provides a systematic framework for decision-making, its implementation has often resulted in unintended consequences (Steen, et al., 2013; Roots, 2004; Lamprianou, 2012). Moreover, researchers question whether the statistical methodologies that underpin EBPM can adequately account for the complexity of educational systems (Bloom, 2019;

Blanchenay, et al., 2014; Wrigley, 2019), which are shaped by cultural, social, and institutional factors that are difficult to quantify (Foster, 2023).

One of the most prominent critiques of EBPM is that it privileges quantifiable indicators of educational success at the expense of broader learning objectives (McKnight & Whitburn, 2020). Policies informed by EBPM generally prioritise measurable performance metrics such as test scores and graduation rates (Wiseman & Davidson, 2018), which reinforces a reductionist view of education that equates learning with numerical outcomes (Geyer, 2012). Critics argue that this focus on statistical indicators and evidence hierarchies has led to policy decisions favouring interventions with demonstrable short-term impacts, even if they do not contribute to long-term educational development (Polesel, et al., 2014; Morsy, et al., 2018).

The reliance on large-scale assessments and quantitative evaluations has also encouraged curriculum narrowing, as schools focus on tested and measured subjects while neglecting less easily quantifiable disciplines (Minarechová, 2012). In many jurisdictions, including Australia, the emphasis on standardised testing regimes such as NAPLAN and PISA has resulted in a decline in instructional time for subjects such as the arts, physical education, and civics (Klenowski & Wyatt-Smith, 2012; Polesel, et al., 2014) as schools prioritise literacy and numeracy performance (Johnston, 2018; Reid, 2020). This phenomenon reflects a broader trend in evidence-based educational policy, where only that which can be measured is valued, leading to the marginalisation of holistic and interdisciplinary approaches to learning (Klenowski & Wyatt-Smith, 2012).

A further limitation of EBPM is its assumption that research findings can be generalised across diverse educational settings (Loughland & Thompson, 2016). Applying statistical models, meta-analyses, and effect size calculations implies that effective interventions in one context will produce similar results in others, if educational systems operate similarly (Cartwright & Hardie, 2012). However, empirical studies have challenged this assumption (Waslander, et al., 2020), demonstrating that the effectiveness of educational interventions is highly dependent on contextual factors (Fullan, 2019) such as teacher expertise, student demographics, and institutional cultures (Sahlberg, 2021).

The risks of policy borrowing, where educational systems adopt strategies from high-performing nations without accounting for local conditions (Herrim, 2014), have been widely documented (Lingard & Sellar, 2013). A notable example is Finland's consistently high performance in PISA (World Population Review, 2024), which has been attributed to teacher autonomy, trust-based governance, and minimal standardised testing (Sahlberg, 2021). However, policymakers in other nations have often sought to replicate Finland's success by implementing isolated policy elements, such as increased teacher training or specific curriculum reforms (Morgan, 2014), without addressing the broader systemic factors that contribute to Finland's educational model (Sellar & Lingard, 2014). Such selective adoption of evidence-based interventions frequently leads to misaligned reforms that fail to deliver expected improvements (Slavin, 2008; Fullan, 2019; Steen, et al., 2013).

Integrating EBPM into educational policy is associated with perverse incentives, particularly in high-stakes accountability systems where funding, school rankings, and teacher evaluations are tied to student performance metrics (Minarechová, 2012). Research has shown that when standardised test scores become the primary measure of educational success, schools and teachers may engage in strategic behaviours to maximise short-term gains, generally at the expense of meaningful learning (Ball, 2012).

2.2.6 Evidence-Based Policymaking in New South Wales

The application of EBPM in education in New South Wales (NSW) has driven reform through frameworks such as the School Excellence Framework (SEF) (NSW Department of Education, 2017) and the School Success Model (SSM) (NSW Department of Education, 2020), both of which emphasise research-backed teaching strategies and performance accountability. The NSW Parliamentary Review (Portfolio Committee No. 3 – Education, 2020) reinforces this approach, drawing substantially on Hattie's (2009) meta-analytic research (Portfolio Committee No. 3 – Education, 2020, p. xii).

However, despite EBPM's prominence, literacy and numeracy outcomes have not consistently improved (Centre for Education Statistics and Evaluation, 2020; Jensen & Kemp, 2011), which has raised concerns about its effectiveness (Biesta, 2010). Critics argue that reliance on standardised metrics enforces rigid, one-size-fits-all policies that limit local adaptability, increase pressure on educators, and narrow the curriculum (Reid, 2020; Polesel, et al., 2014). Furthermore, implementation challenges arise from a disconnect between policymakers and practitioners (Johnston, 2018), with teachers frequently expressing frustration at data-driven mandates that fail to acknowledge classroom realities (Fuller, 2022).

The NSW experience underscores both the strengths and limitations of EBPM and highlights the risks of data-driven policymaking detached from contextual realities (Hardy, 2013). As a result, scholars increasingly advocate for hybrid models in education, such as evidence-based policy and practice (Brown, 2018), which integrate empirical data with qualitative insights, practitioner experience, and community engagement (Ansell & Geyer, 2017).

2.2.7 Evidence-Based Policymaking: A double-edged sword

EBPM has fundamentally reshaped how educational policy is formulated, evaluated, and implemented, providing a systematic, data-driven approach to decision-making (Bromley, et al., 2023). Integrating empirical research, meta-analysis, and statistical modelling into education governance has enabled policymakers identify potentially high-impact interventions (Hattie, 2023), which foster greater accountability (Geyer, 2012), transparency, and efficiency in policy design. The expansion of ISLA regimes, such as PISA and TIMSS, has reinforced the global diffusion of EBPM and encouraged national educational systems to adopt standardised performance measures and evidence-based improvement strategies (Breakspear, 2012).

EBPM remains a contested (Reid, 2020) and evolving (Ansell & Geyer, 2017) paradigm, with substantial critiques emerging regarding its methodological limitations (O'Neill, 2012), its over-reliance on quantifiable metrics (Brown, 2018), and its failure to account for the complex, context-dependent nature of education (Gorard, 2020). While meta-analysis and regression-based statistical models have promised insights into the effectiveness of different interventions (Parker, et al., 2018), they have also been criticised for oversimplifying educational phenomena and assuming generalisability (Fullan, 2019; Waslander, et al., 2020). The limitations of EBPM (Christensen & Laegreid, 2022) emphasise the need for greater pluralism in educational research and policy (British Educational Research Association, 2017), recognising the methodological difficulty in addressing the complexities of learning and teaching (Wiseman & Davidson, 2018). Finally, the associated rise of audit and compliance has led to policy volatility (Geyer, 2012), which raises concerns about the long-term sustainability of evidence-based reforms (Reid, 2020).

2.3 The contribution of infrastructure in the context of Evidence-Based Policymaking

The relationship between infrastructure and educational outcomes continues to be a topic of considerable study, with researchers drawing differing conclusions that are often shaped by their chosen methodology. While some studies identify infrastructure as having limited influence on student achievement (Belmonte, et al., 2019; Martorell, et al., 2016), others highlight its more significant role (Berman, et al., 2018; Hong & Zimmer, 2016), particularly its influence on other interventions (Crampton, 2009).

The evaluation of the impact of infrastructure has been strongly influenced by the principles underpinning EBPM, which prioritises empirical, quantifiable data as the basis for policy decisions (Christensen & Laegreid, 2022; Cartwright & Hardie, 2012). Within the EBPM framework, researchers isolate specific variables, such as infrastructure quality (Gunter & Shao, 2016) or spending (Belmonte, et al., 2019), to assess their direct impact on measurable outcomes such as academic achievement. This reductionist approach has contributed to a de-emphasis on infrastructure in educational policy (Buckley, et al., 2005), where meta-analytic findings suggest that its independent effect is minimal (Hattie, 2023; Martorell, et al., 2016).

However, an alternative view emerges from the research that adopts a systems-based perspective, which positions infrastructure as an enabling factor supporting the success of other high-impact interventions (Crampton, 2009). This section of the literature review examines how EBPM's methodological constraints have shaped infrastructure research; identifies key limitations in isolating infrastructure as an independent factor; and explores how a more integrated approach could provide a fuller understanding of its educational significance (Maxwell, 2016).

2.3.1 Overall infrastructure condition and expenditure

Much of the literature examining the contribution of infrastructure to educational outcomes uses infrastructure spend as the independent variable (Hong & Zimmer, 2016). Studies have found a small positive correlation between increased spend and improved educational outcomes (Gunter & Shao, 2016). For example, Belmonte et al. (2019) found increased infrastructure spend correlated with a small improvement in educational outcomes when comparing results achieved by schools following a regional rebuilding program. Likewise, Crampton (2009) found a statistically significant relationship between increased infrastructure expenditure and improved educational outcomes. However, when compared to other factors, Hattie (2023) found the potential for infrastructure to improve educational outcomes was insignificant.

Gunter and Shao's (2016) meta-analysis concluded that school building conditions were only weakly correlated with academic achievement. However, the authors cautioned that this finding should not be interpreted as evidence that infrastructure plays an insignificant role in education. Instead, they suggested that such weak correlations often arise from the difficulty of capturing the indirect effects of infrastructure within traditional quantitative frameworks. Significantly, Roberts (2009) found that the presence of a relationship between infrastructure condition and functionality depended on the assessor being a school user rather than an external expert. The importance of user experience was also noted by Buckley et al. (2005), who observed a significant relationship between teacher satisfaction with infrastructure and teacher turnover.

2.3.2 Classroom environmental conditions

Empirical studies have identified lighting, temperature, noise levels, and air quality as significant determinants of cognitive performance and educational outcomes. Lighting has been found to play a critical role in supporting learning. Toyinbo et al. (2024) reported that classrooms equipped with daylight-mimicking LED systems improved reading comprehension by up to 15%. Similarly, Heschong et al. (2002) found that students in well-lit classrooms outperformed their peers by up to 26% in reading and 20% in mathematics. Inadequate lighting, by contrast, has been shown to increase visual strain and reduce attentiveness, thereby undermining student performance.

Research has also demonstrated the adverse effects of noise pollution, which impedes auditory processing and raises stress levels. Al-Khatri et al. (2024) found that noise reduction strategies, such as installing acoustic insulation, improved test performance by approximately 20%. Similarly, temperature regulation has been shown to influence academic performance, with optimal conditions improving focus and reducing fatigue (Wargocki & Wyon, 2013). Poor air quality has also been linked to higher absenteeism rates and lower cognitive performance, with improved ventilation systems yielding substantial gains in student achievement (Totfum, et al., 2015). These findings illustrate the potential value of standard statistical processes aligned to EBPM when conditions are constrained, and factors are reasonably separable and measurable.

2.3.3 The potential for Evidence-Based Policymaking to overlook infrastructure

Although EBPM-driven evaluations frequently conclude that infrastructure has a minimal direct effect on student achievement (Martorell, et al., 2016), an alternative body of research emphasises its role as an indirect contributor to educational outcomes (Uline & Tschannen-Moran, 2008; Maxwell, 2016). Studies adopting a broader perspective suggest that the contribution of infrastructure to educational outcomes is likely to be indirect (Wang & Degol, 2016), and potentially more significant than is currently accepted (Crampton, 2009).

Research by Buckley et al. (2005) demonstrated that improved school facilities enhance teacher satisfaction and retention, which have well-documented links to improved student engagement (Jiang, et al., 2022). Similarly, Urick and Bowers (2014) and Berman et al. (2018) identified an association between well-maintained school environments and improved student attendance and behaviour, factors that indirectly support educational success. Further evidence of interdependencies between inputs in education was highlighted by Crampton (2009), who found that the effect of infrastructure spend was influenced by spend on other, non-infrastructure factors. Further, Crampton noted that the effect of non-infrastructure spend was improved by coincident infrastructure spend.

EBPM has profoundly shaped educational reform (Bromley, et al., 2023; Sahlberg, 2007) by prioritising interventions with demonstrable, quantifiable impacts on student achievement (Geyer, 2012; Chezan, et al., 2017). Infrastructure, excluding indoor environmental conditions (Wargocki & Wyon, 2013), has often been positioned within this framework as a secondary concern, potentially because meta-analytic methodologies and statistical models have struggled to capture its indirect (Crampton, 2009) and context-dependent contributions to learning outcomes (Goss, 2024). However, a growing body of research suggests that infrastructure should not be viewed as a discrete variable but as an integral component of a broader ecosystem (Rudasill, et al., 2018) that enables effective teaching and learning (Berman, et al., 2018).

High-quality school facilities are shown to contribute to student wellbeing, attendance, and cognitive performance (Berman, et al., 2018) while also fostering teacher satisfaction (Buckley, et al., 2005) and retention—factors that, in turn, enhance educational outcomes (Maxwell, 2016). The prevailing reductionist tendencies of EBPM, which seek to isolate causal relationships, have potentially obscured the complex, mediating role of infrastructure in the learning environment and subsequent educational outcomes.

2.4 School Climate: The role of experience in shaping outcomes

Section 2.3.3 examined the contribution of infrastructure to education as viewed through the lens of EBPM, which seeks to identify, isolate, and measure direct causal relationships between educational inputs and student outcomes (Ansell & Geyer, 2017). Using this approach, policymakers have often prioritised interventions with the highest measurable impact on standardised test performance (Blass, 2020), leading to the de-prioritisation of infrastructure investment (Crampton, 2009; Holmund, et al., 2010) in favour of interventions deemed more effective by virtue of a greater effect size (Simpson, 2017). Using this and similar approaches, researchers have found the influence of infrastructure to be small (Gunter & Shao, 2016), difficult to define beyond expenditure (Belmonte, et al., 2019), and insufficient for consideration in the development of improvement policy (Hattie, 2023; Martorell, et al., 2016). However, an evolving body of research offers a different perspective on how School Climate, a method of describing school experience, shapes educational outcomes (Thapa, et al., 2013). Increasingly, researchers are considering infrastructure as influencing or contributing to School Climate (Wang & Degol, 2016; Maxwell, 2016).

Rather than attempting to quantify the independent, direct impact of infrastructure, researchers in the field of School Climate argue that infrastructure plays a critical enabling role, shaping the conditions that support, or hinder, effective teaching and learning (Uline & Tschannen-Moran, 2008). School Climate research emphasises the interconnected nature of educational inputs (Rudasill, et al., 2018), suggesting that a school's physical environment interacts with teacher morale, student engagement, and school culture to influence learning outcomes (Thapa, et al., 2013; Daily, et al., 2019). This section explores the School Climate framework as a means of linking educational inputs and outcomes (Wang & Degol, 2016) and how it may provide an integrated and context-sensitive approach to understanding the relationship between infrastructure and educational effectiveness.

2.4.1 School Climate: An evolving construct

School Climate research provides an alternative framework for assessing the effectiveness of educational environments, shifting the focus from isolated inputs to the overall experience of students, teachers, and parents (Zullig, et al., 2010). Wang and Degol (2016) describe School

Climate as encompassing "virtually every aspect of the school experience" (p. 315). School Climate has been shown by researchers to be predictive of behaviour (Benbenishty, et al., 2016), wellbeing (Berman, et al., 2018), and academic performance (Daily, et al., 2019). School Climate has also been shown to be a mitigator of poverty on academic and behavioural outcomes (Hopson & Lee, 2011).

Following the initial work of Halpin and Croft (1963) in defining School Climate, researchers developed the underlying framework to the point where three main pillars together described the experience of the school (Zullig, et al., 2010): Safety, Community (referred to here as Interpersonal Relationships), and Academic Press (referred to here as Teaching and Learning) (Thapa, et al., 2013). However, in reviewing the literature, Wang and Degol (2016) noted numerous studies included the effect of the physical environment on School Climate. As such, they proposed that School Climate is composed of four pillars: Safety, Community (Interpersonal Relationships), Academic Press (Teaching and Learning), and Institutional Environment. These four pillars can be broken down further into more detailed constructs, as shown in Table 2.1. Of specific interest is the environmental construct within the Institutional Environment pillar as this details the features of user experience with infrastructure.

Table 2.1. The conceptualisation and categorisation of School Climate (Wang & Degol, 2016, p. 318)

Pillar	Construct	Salient features
Safety	Social/emotional	Lack of bullying counselling
	Discipline & order	Conflict resolution, clarity, fairness, and consistency of rules, belief in school rules
	Physical	Reduced violence and aggression, students and staff feel safe, security measures
Community (Interpersonal Relationships)	Partnership	The role that community members and parents play, parental involvement
	Quality of relationships	Trust, interpersonal relationships between staff and students, affiliation
	Connectedness	Cohesion, sense of belonging, student activities
	Respect for diversity	Fairness, autonomy, opportunities for decision-making (teachers and students), cultural awareness
Academic Press (Teaching and Learning)	Leadership	Principals and administration supportive of teachers, open lines of communication
	Instruction	Quality of instruction, assessment of students, willingness of teacher, motivation of student, teacher expectations, achievement goal structure, teachers' use of supportive practices
	Professional development	Review and assessment of teaching practices, opportunities for growth and professional development
Institutional Environment	Environmental	Heating, lighting, air-conditioning, acoustical control, cleanliness, upkeep of maintenance, quality of construction and fit-out
	Structural organisation	Class size, student-to-teacher ratio, school size, ability tracking

	Availability of resources	Adequacy of supplies, resources and materials, technology, sharing of resources

The School Climate framework, and the dimensions describing it, continue to change and evolve. Some researchers have adopted the idea of Institutional Environment but excluded infrastructure as part of this (Aldridge, et al., 2024), whereas others treat infrastructure as separate to School Climate (Uline & Tschannen-Moran, 2008; Goss, 2024). Despite growing acknowledgement that infrastructure plays a role in School Climate (National School Climate Center, 2021), there exists little evidence of the integration of infrastructure into instruments measuring School Climate.

The School Climate construct continues to change and evolve, with the National School Climate Center (2021) recently adding Leadership and Efficacy to the four pillars proposed by Wang and Degol (2016). While the ongoing development of the School Climate framework reflects its continued relevance and ongoing research (Aldridge, et al., 2024; Goss, 2024), it also reinforces the concerns of researchers who note the difficulty of comparing the findings of studies where the framework is neither standardised nor static (Thapa, et al., 2013; Wang & Degol, 2016).

2.4.2 School Climate: Repositioning the contribution of infrastructure

A key distinction between the methods commonly applied in EBPM and School Climate is how they conceptualise the contribution of infrastructure (and other inputs) to educational outcomes. Traditional approaches evaluate infrastructure in terms of its direct effect on student achievement, generally using meta-analysis (Martorell, et al., 2016) and regression techniques (Belmonte, et al., 2019) to compare its impact against other interventions. The input measure used is typically expenditure, as this is readily available and comparable (Crampton, 2009; Gunter & Shao, 2016).

In contrast, School Climate research incorporates the experience of infrastructure within a broader assessment of school experience (Wang & Degol, 2016). By framing infrastructure as a component of school experience, School Climate research allows the consideration of indirect effects and interactions with other factors (Maxwell, 2016), defined as "interplay" by Uline and Tschannen-Moran (2008). The potential for methods associated with EBPM to overlook the contribution of infrastructure was noted by Gunter and Shao (2016). Similarly, Crampton (2009) found that the effectiveness of non-infrastructure interventions was increased if undertaken with coincident infrastructure investment.

The literature remains unclear on how infrastructure influences educational outcomes; however, researchers have posited several mechanisms (Gislason, 2009; Goss, 2024). High-quality infrastructure, for example, may enhance student concentration (Wargocki & Wyon, 2013), reduce teacher turnover (Buckley, et al., 2005), and create a sense of school pride, all of which contribute to better learning outcomes (Goss, 2024). Conversely, poor infrastructure may contribute to a negative school climate, leading to increased absenteeism (Maxwell, 2016), reduced engagement, and lower morale among staff and students (Urick & Bowers, 2014). This broader conceptualisation helps explain why traditional statistical analyses may not fully capture the true impact of infrastructure on educational outcomes. Techniques traditionally applied in the pursuit of improved educational outcomes are potentially unsuitable when the relationships between inputs and outcomes are indirect (Fullan, 2019), interdependent (Loughland & Thompson, 2016), and context-specific (Berman, et al., 2018) rather than immediate and quantifiable in isolation (Simpson, 2018).

2.4.3 School Climate: Criticisms of the construct

Despite its contributions, the School Climate framework has limitations that challenge its conceptual and methodological robustness (Zullig, et al., 2010). Scholars identified issues related to its lack of standardised definitions, reliance on subjective data, and difficulties in policy application (Konold, et al., 2018).

One major critique of School Climate is the absence of a single, universally accepted definition of School Climate (Wang & Degol, 2016). This lack of a standard construct leads to variability in research frameworks and findings (Van Houtte & Van Maele, 2011), and difficulties translating School Climate into policy frameworks (Thapa, et al., 2013). Additionally, the reliance of School Climate on perceptual surveys introduces bias and subjectivity, since students, teachers, and administrators may interpret school environments differently (Zabek, et al., 2022).

Studies linking infrastructure and School Climate commonly rely on qualitative analysis (Goss, 2024; Gislason, 2009), which potentially limits the predictive utility for policymakers. Moreover, critics argue that quantitative analysis used in School Climate studies often establishes correlation rather than causation, making it unclear whether a positive school climate directly improves student outcomes or merely reflects broader structural advantages (Konold, et al., 2018). This last point is interesting because it highlights a potential inconsistency woven through the quantitative School Climate research. Specifically, while School Climate emphasises interconnectedness of different elements as a strength (Rudasill, et al., 2018; Hanuliaková & Barnová, 2015), researchers generally rely on regression analysis to isolate and quantify the effect of individual elements (Uline & Tschannen-Moran, 2008; Benbenishty, et al., 2016). This approach risks oversimplifying or overlooking the same interconnections that School Climate theory seeks to highlight (Feldhoff, et al., 2021) and that are central to the claimed advantages of School Climate (Zabek, et al., 2022) when compared to more singular approaches and their potential limitations (Bloom, 2019).

2.5 Schools as complex and dynamic systems

Educational systems are intricate networks where numerous interconnected elements (Rudasill, et al., 2018)—including students, teachers, administrators, policies, and community factors—interact dynamically and often unpredictably (Koopmans & Stamovlasis, 2016). Traditional policy approaches have frequently operated under predominantly linear assumptions (Geyer, 2012; Reid, 2020), suggesting that specific inputs will lead to predictable outcomes (Lingard, 2013). However, the phenomenon of policy resistance, where well-intentioned interventions yield unexpected or counterproductive results (Waslander, et al., 2020), challenges this linear perspective (Blanchenay, et al., 2014).

Despite numerous jurisdictions implementing evidence-based policies (Putansu, 2020), anticipated improvements in educational outcomes have not always materialised (Morrison, 2021; Reid, 2020). In examining the difficulty in translating statistically significant relationships into successful improvement policy, Bloom (2019) noted the misalignment between the nature of schools and the assumptions inherent in approached influencing EBPM. Mason (2016), like Steen et al. (2013), also noted the incompatibility between expectations of policy-led improvement and the complex reality of schools. As such, the growing awareness of complexity as a factor prompts an understanding and consideration of complexity theory (Mitchell, 2009), particularly in the context of educational policy (Mason, 2008), and the role of infrastructure therein.

2.5.1 Complexity theory: Understanding dynamic systems

Complexity theory offers a lens through which to examine systems characterised by numerous interconnected components whose interactions yield emergent, often unpredictable behaviours (Mitchell, 2009). Unlike complicated systems, where elements function in a predictable, linear manner, complex systems exhibit properties such as emergence, feedback loops, non-linearity, and self-organisation (Koopmans & Stamovlasis, 2016).

Emergence refers to system-wide patterns arising from local interactions among components, leading to behaviours not easily inferred from the individual parts (Waslander, et al., 2016). Non-linearity implies that small changes can have a disproportionately large impact, which makes outcomes highly sensitive to initial conditions (Spataru, 2015). Feedback loops occur when interconnected parts affect each other and serve to either reinforce or balance system behaviour (Sterman, 2000). Self-organisation denotes the capacity of systems to evolve and adapt without centralised control, responding to internal and external stimuli in dynamic ways (Steen, et al., 2013). These characteristics are prevalent in social systems, including education (Rudasill, et al., 2018), where interactions among individuals and institutions result in complex adaptive behaviours (Morrison, 2010).

The theoretical origins of complexity theory lie in diverse disciplines including physics, biology, cybernetics, and systems theory. Prigogine and Stengers (1984) introduced the concept of dissipative structures, which are systems that maintain order through exchanges with their environment, even under conditions of instability. Capra (1996) advanced this view, emphasising interdependence and systemic coherence in living systems. Holland (1995) and Gell-Mann (1994) further developed the concept of complex adaptive systems by exploring how agents within such systems learn, adapt, and evolve through interactions governed by simple rules.

Despite its conceptual richness, complexity theory is not without limitations. Critics have highlighted its theoretical ambiguity, particularly the inconsistent use of core terms such as "emergence" and "self-organisation", which may be applied metaphorically rather than analytically (Davis & Sumara, 2006). The rejection of linear causality and prediction complicates the development of systematic approaches to modelling or understanding complex systems (Vulic, et al., 2024). Moreover, while complexity theory foregrounds decentralised processes and adaptive change, these ideas present practical difficulties, particularly when used to design or contribute to policy that is to be implemented (Mason, 2016). Recent literature has sought to refine and clarify the theoretical core of complexity theory. For example, Yang (2024) revisited foundational assumptions to distinguish between complex and merely complicated systems more precisely while challenging the use of linear methods to support analysis derived from non-linear methods. Such efforts signal the ongoing evolution of complexity theory as a conceptual tool for grappling with uncertainty, interdependence, and emergence across a wide range of domains.

2.5.2 The case for treating schools as complex and dynamic systems

This section of the literature review examines the idea that schools are likely to exhibit dynamic complexity, which is a contributor to the differences between the expectations and reality of EBPM in education (Biesta, 2010; Geyer, 2012; Snyder, 2013). Essentially the dominant methods used in the development of educational policy assume schools are complicated (Goldspink, 2007).

Schools epitomise complex systems due to the intricate interdependencies between teachers, students, administrators, and the broader community (Rudasill, et al., 2018). These interrelations contribute to emergent behaviours that defy simple prediction (Morrison, 2010; Koopmans, 2020). For example, identical policy interventions can lead to divergent outcomes across different schools (Waslander, et al., 2020), as demonstrated by Steen et al.'s (2013) comparative case studies of schools exhibiting varying responses to the same improvement strategies.

In conducting a post-mortem evaluation of a school improvement program in the Netherlands, Steen et al. (2013) noted that policymakers had assumed “linear, proportionate causation between interventions and consequences” (p. 551). The objective of this study was to understand the different outcomes following the application of educational policy to schools deemed to be similar. Steen studied four different schools and concluded that the cause of the different outcomes was the emergence of different dynamics. The cause of these dynamics was competing and reinforcing loops, which are the building blocks of complex dynamic systems. The following discussion presents two of the four case studies from Steen et al. (2013) to illustrate the different results obtained from the same intervention at two similarly underperforming schools.

In Figure 2.1, Steen et al. (2013, p. 562) used a causal loop diagram, a precursor to Systems Dynamics Modelling (SDM) (Sterman, 2000) to demonstrate School A’s reaction to the improvement policy. A high-stakes assessment deemed the school to be “very weak”. The education department responded by sending a team from the Quality Assessment Inspectorate to provide more detailed oversight and identify specific improvement actions.

School A

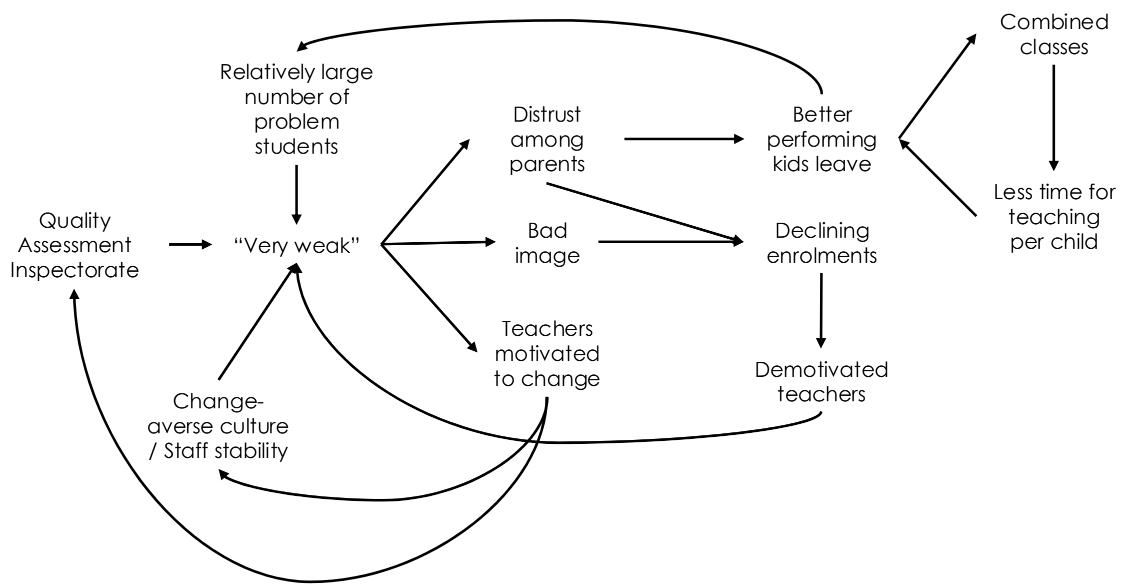

Figure 2.1 Causal loop diagram illustrating an unsuccessful policy intervention in a “very weak” school (Steen et al., 2013, p. 562)

Classifying the school as “very weak” resulted in better-performing students leaving the school and teachers becoming increasingly demotivated. Steen et al. (2013) viewed these two reactions in combination to be the cause of the subsequent decline. The reduction in high-performing students increased the proportion of problem and underperforming students. Demotivated teachers and other teachers resistant to change reinforced the perception of a "very weak" school. These two (negative reinforcing) loops accelerated the school's decline to the point where it was closed.

School B (Figure 2.1) was also deemed "very weak" and the school improvement policy subsequently implemented.

School B

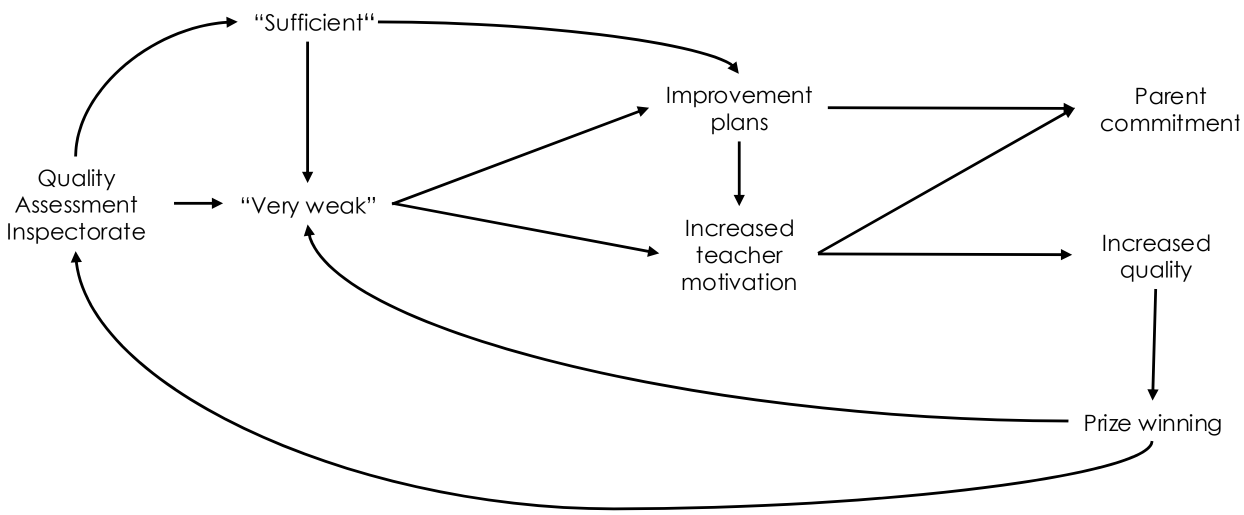

Figure 2.2 Causal loop diagram illustrating a successful policy intervention in a "very weak" school (Steen et al., 2013, p. 562)

Following the intervention of the Quality Assessment Inspectorate, School B produced improvement plans that motivated both teachers and parents. In this case, the positive reinforcing loop accelerated the improvement of School B, which continued after the inspectorate left (Steen, et al., 2013).

The work of Steen et al. (2013), together with that of Waslander et al. (2020) and Venter (2022), supports the idea that schools operate as complex and dynamic systems. Using causal loop diagrams, these researchers illustrated how a complexity-informed approach can explain the divergent outcomes observed when ostensibly similar interventions are applied across different schools (Steen, et al., 2013) and within broader educational systems (Waslander, et al., 2020). Their findings emphasise that system behaviours emerge from multiple context-dependent factors (Rudasill, et al., 2018), making them resistant to simplistic, linear analysis and intervention. The idea that schools exhibit complex and dynamic behaviour challenges the assumptions underpinning traditional approaches to educational reform, particularly those aligned with EBPM (British Educational Research Association, 2017). Linear models of change, which presuppose predictable and controllable cause-and-effect relationships, have consistently proven inadequate in environments characterised by feedback loops, emergence, and sensitivity to initial conditions (Head, 2016; Sanderson, 2009). Nonetheless, EBPM continues to dominate educational policy discourse (Reid, 2020; Blass, 2020; Bromley, et al., 2023). As Ansell and Geyer (2017) argue, this persistence may reflect that it is less about the empirical success of EBPM in complex contexts than the current absence of a widely accepted alternative framework capable of accommodating the demands of complexity in public policy and practice.

2.5.3 The mathematical challenge of complexity

Mathematically accounting for complexity in schools presents significant challenges (Tin, et al., 2024). Traditional statistical techniques, particularly regression analysis, rely on assumptions that are often violated in educational systems (Blass, 2020). These methods presuppose stable, linear, and independent relationships between variables (Bloom, 2019); however, complex systems frequently exhibit non-linearity and interdependence (Morrison, 2010).

A fundamental mathematical limitation of regression analysis is its assumption of linearity (Foster, 2023). These models typically represent relationships as additive and proportional, meaning that a change in an independent variable leads to a consistent and predictable change in the dependent variable (Donnelly & Abdel-Raouf, 2016). However, many real-world systems display non-linear behaviour (Mitchell, 2009), where effects may be contingent on thresholds, interaction terms, or diminishing returns (Benkhalfallah & Laouar, 2023). Standard linear models fail to capture the underlying mathematical structure in such cases, which can lead to inaccurate estimations and misinterpretation of causal relationships (Waslander, et al., 2020).

Another core issue is the assumption of independence among observations. Many statistical techniques require each data point to be independently drawn from the same distribution (Donnelly & Abdel-Raouf, 2016). However, variables often exhibit interdependencies in complex systems through structured relationships (Rudasill, et al., 2018), spatial correlations, or network effects (van Vondel, et al., 2016). When such dependencies exist, conventional models tend to underestimate standard errors, inflate significance levels, and misrepresent the actual structure of the data (Benkhalfallah & Laouar, 2023).

Given these mathematical challenges, conventional statistical methods frequently struggle to adequately represent complex systems. Alternative approaches such as systems modelling, network analysis, and machine learning offer more robust mathematical frameworks by accommodating non-linear relationships, interdependent structures, and emergent patterns (Larose & Larose, 2015).

2.5.4 Relevant computational advances

Computational modelling offers promising alternatives for capturing the complexity of educational systems and addressing the inadequacies of traditional statistical approaches. Systems Dynamics Modelling (SDM) simulates the long-term effects of policy decisions by modelling the interdependencies and feedback loops within complex and dynamic systems (Warren, 2015). While successfully applied in healthcare and environmental policy (Spataru, 2015), SDM application in the educational context remains underexplored (van Vondel, et al., 2016). Causal loop diagrams, a precursor to SDM, have been used in the educational context by Steen et al. (2013) and Waslander et al. (2016); however, in both cases, causal loop diagrams were employed to map system responses to past interventions rather than predict responses to potential future interventions.

Machine learning as applied to the analysis of complex and dynamic systems uses artificial intelligence techniques to uncover hidden patterns in data, offering insights beyond the reach of traditional methods (Wu & Coggeshall, 2012). However, the opacity of these models has traditionally presented challenges for interpretability (Zaman, et al., 2023; Delgado-Panadero, et al., 2022) and while potentially promising, researchers note the lack of application in education (Wang & Degol, 2016). As Baker et el. (2016) observe, most machine learning applications in educational research have concentrated on classification or prediction tasks that are focused on singular, narrowly defined outcome measures such as examination scores, attendance, or course completions. While useful in context, these approaches typically treat educational outcomes as isolated variables and fail to account for the interdependencies and multidimensional nature of educational achievement (Romero & Ventura, 2010).

Agent-based modelling (ABM) has been proposed as a viable approach for addressing the complexities inherent in educational systems by simulating the localised interactions among individual agents within learning environments (Kazakov, et al., 2024). This methodological framework provides valuable insights into emergent behaviours that arise from these interactions; however, its practical application within the field of education remains generally theoretical. Persistent challenges related to empirical validation and the integration of diverse data sources continue to hinder the effectiveness of ABM in educational contexts (Koster, et al., 2017; Gu & Blackmore, 2015; Vulic, et al., 2024).

Although the methodological potential of machine learning to engage with the dynamic and non-linear characteristics of educational systems is considerable (Zawacki-Richter, et al., 2019), there remains a relative lack of research explicitly addressing the prediction of holistic, integrated outcome patterns that better reflect the complexity of real-world educational experiences.

2.5.5 Challenges facing the application of complexity theory in the educational context

While complexity theory provides a compelling framework for understanding the dynamic and interconnected nature of educational systems (Steen, et al., 2013), its practical application in policy settings presents significant challenges. One of the foremost difficulties lies in translating complexity-based insights into actionable policy measures (Ansell & Geyer, 2017). Complexity theory acknowledges that schools operate as adaptive systems with emergent behaviours (Koopmans, 2020), but this does not readily translate into clear, prescriptive guidelines for policymakers seeking to implement effective reforms (Tin, et al., 2024). Educational policy requires concrete, scalable solutions (Bromley, et al., 2023), whereas complexity-based approaches often emphasise local adaptation and context-dependent strategies that may resist standardisation (Johnson, 2008).

Another challenge is the need for policymakers to balance flexibility with accountability (Brown, 2018). Traditional educational policy relies on measurable indicators, such as standardised test scores and graduation rates, to assess the effectiveness of interventions (Wiseman & Davidson, 2018). Complexity theory, however, suggests that educational success emerges from non-linear and context-dependent interactions, which makes it difficult to establish universal benchmarks for success (Spataru, 2015). As a result, tension can arise between the need for systematic evaluation and the inherent unpredictability of complex systems (Feldhoff, et al., 2021), which is described by Bloom (2019) as a "clash of paradigms and conflicting assumptions" (p. 24). Further, policymakers are generally reluctant to embrace models that do not provide transparent cause-and-effect relationships (Reid, 2020), mainly when political and public expectations demand demonstrable results (Putansu, 2020). Implementing complexity-based methodologies, such as SDM and ABM, requires advanced technical expertise and computational resources not always available within education departments.

2.5.6 Towards a complexity-informed approach to educational policy and infrastructure

Recent literature increasingly conceptualises schools as complex adaptive systems (Koopmans, 2020), a perspective that challenges the assumptions underpinning linear policy models. These models typically propose that policymakers can achieve predictable outcomes by manipulating discrete inputs. However, this approach may not fully account for the non-linear, interdependent dynamics that characterise educational environments (Bromley, et al., 2023). Geyer (2012) underscores the limitations of conventional statistical techniques in examining the behaviour of complex and dynamic systems, as they often overlook key system features such as recursive feedback loops, emergent behaviours, and adaptive responses to context.

Researchers have responded to these challenges by adopting computational approaches, notably, SDM and ABM, which provide more accurate representations of dynamic educational systems (Farhood, et al., 2024). These methods enable policymakers and analysts to explore how various system components interact over time. However, the effective integration of such tools into policy frameworks requires caution. Brown (2018) argues that although these methods may enable nuanced insights, their real-world utility depends on their alignment with policymaking logic, which generally favours scalable and measurable solutions over complex modelling outputs.

Complexity theory proves potentially valuable when applied to the domain of school infrastructure, which policymakers have traditionally treated as a discrete variable. This reductive view isolates infrastructure from broader pedagogical and contextual factors, thereby potentially underestimating its systemic role. In contrast, complexity-informed perspectives position infrastructure as a relational element within a dynamic ecosystem, continuously interacting with other components of the school environment (Rudasill, et al., 2018). Scholars such as Uline and Tschannen-Moran (2008) have demonstrated that the built environment influences educational outcomes by influencing conditions such as school climate, teacher satisfaction, and student wellbeing. Evidence increasingly supports this systemic view, with Buckley et al. (2005) demonstrating that investment in infrastructure contributes to teacher retention. Likewise, Urick and Bowers (2014) link well-maintained physical environments to enhanced perceptions of safety and student engagement, which are factors that, while not directly academic, strongly influence learning outcomes. These findings highlight the epistemic limitations of dominant evaluation models, which tend to privilege what is observable and measurable over what is experientially meaningful.

Despite such developments, many national and local policies rely on narrow definitions of evidence, typically inherited EBPM (Bromley, et al., 2023). This approach may obscure the complex interplay between infrastructure, pedagogy, and learner experience in educational settings (Wang & Degol, 2016). Scholars have, therefore, advocated for a broader epistemological framework that values contextual relevance, practical utility, and the non-linear nature of educational change (Jørgensen, 2024; Newman, et al., 2017). Oliver (2022) observes that although researchers are increasingly recognising the importance of context and complexity, many continue to rely on rationalist assumptions and technocratic preferences when formulating policy. Bloom (2019) makes a similar argument, suggesting that schools, as dynamic systems, require analytical tools that transcend linear causality. EBPM risks oversimplifying reform efforts without such conceptual shifts (Mason, 2016) by prioritising what is easy to measure over what is educationally significant (Loughland & Thompson, 2016).

A growing body of research now supports a more integrated understanding of infrastructure as a policy lever. For example, Crampton et al. (2009) found that when combined with instructional reforms, infrastructure investment can lead to increased gains in student outcomes. This suggests that infrastructure is a key enabler that creates the conditions necessary for other reforms to take hold (Maxwell, 2016).

In summary, complexity theory does not offer prescriptive solutions but provides a conceptual framework for navigating the uncertainties, interdependencies, and contextual specificities of educational policy (Steen, et al., 2013). By recognising the indirect and interactive effects of infrastructure and challenging hierarchical models of evidence, policymakers can develop more resilient, responsive, and context-sensitive approaches to reform (Ansell & Geyer, 2017; Koopmans, 2020). This complexity-informed orientation underpins the literature synthesis that follows.

2.5.7 Synthesising the evidence linking infrastructure to educational outcomes

This literature review highlights the tensions in how the role of infrastructure in educational outcomes is conceptualised, measured, and valued (Reid, 2020; British Educational Research Association, 2017). Traditional evidence-based approaches, influenced by the priorities of ILSAs and the Global Education Reform Movement (GERM), tend to emphasise interventions that produce immediate, quantifiable gains in standardised assessments (Bromley, et al., 2023; Sahlberg, 2007). Within this framework, infrastructure is often marginalised and regarded primarily as a background condition rather than an active contributor to student learning (Gunter & Shao, 2016). Empirical studies reveal limitations within this framing (Crampton, 2009; Maxwell, 2016). Regression-based analyses commonly identify small but statistically significant associations between infrastructure investment and student achievement (Martorell, et al., 2016). However, they frequently rely on financial expenditure data as a proxy for improvements in the learning environment (Hong & Zimmer, 2016). Meta-analyses, such as Hattie's (2023) synthesis, have reinforced the perception of the limited direct impact of infrastructure; however, the aggregation of studies from diverse contexts risks obscuring more subtle or indirect effects.

In contrast, research adopting systems-based perspectives (Rudasill, et al., 2018), such as the School Climate framework (Thapa, et al., 2013), presents a more complex understanding of the influence of infrastructure. Studies suggest that infrastructure affects educational outcomes indirectly by shaping factors such as teacher satisfaction, student engagement, and school climate (Urick & Bowers, 2014; Buckley, et al., 2005). Research by Roberts (2009) further indicates that perceived infrastructure quality, as reported by students and teachers, correlates more strongly with achievement than externally assessed facility conditions. These findings highlight the importance of user experience as a mediating factor in the relationship between infrastructure and educational outcomes (Maxwell, 2016).

Insights from complexity theory reinforce this broader conceptualisation. The literature increasingly characterises schools as dynamic systems (van Vondel, et al., 2016; Koopmans, 2020) in which multiple factors—such as infrastructure, pedagogy, leadership, and community engagement—interact non-linearly and context-dependently (Steen, et al., 2013). Policy resistance and adaptive systems studies suggest that identical interventions can yield divergent outcomes depending on the local school context (Waslander, et al., 2020), challenging assumptions underpinning conventional linear educational improvement models.

Three areas of limitation emerge from the existing literature. First, research frequently adopts a narrow conceptualisation of infrastructure, which focuses on objective physical conditions (Gunter & Shao, 2016) or financial investments (Hong & Zimmer, 2016) while overlooking its broader role in shaping the educational environment (Maxwell, 2016). Second, most studies employ analytical models that treat educational factors as independent variables rather than examining their complex interdependencies (Crampton, 2009). Third, although machine learning and systems modelling offer the potential for capturing non-linear interactions and emergent dynamics, their application to infrastructure research remains limited (Zaman, et al., 2023; Zawacki-Richter, et al., 2019).

These findings suggest a growing recognition within the literature that infrastructure may act not simply as an isolated input but as a contextual factor influencing the effectiveness of broader educational strategies. Recent studies propose that more integrative methodologies, drawing on complexity theory, School Climate research, and advanced analytical techniques, may be necessary to capture the full contribution of infrastructure to educational outcomes. This emerging perspective provides the foundation for formulating this study's research question.

2.6 Deriving the research question

The review of existing research exposes two fundamental limitations in how educational infrastructure is conceptualised and investigated. First, most studies adopt linear, input–output models grounded in EBPM paradigms. These models typically operationalise infrastructure as a quantifiable variable, often financial or physical, and apply inferential statistical methods, particularly regression analysis, which assume stable, independent causal relationships (Koopmans, 2020; Geyer, 2012). Such approaches oversimplify the educational process by treating schools as static rather than adaptive systems and frequently overlook the complex interactions between environmental, organisational, and individual variables (Biesta, 2010; Rudasill, et al., 2018).

Second, while empirical studies consistently report associations between improved infrastructure (such as lighting, air quality, and acoustics) and student outcomes, these relationships are rarely theorised in a way that explains how or why such effects occur (Hong & Zimmer, 2016; Shield, et al., 2010; Totfum, et al., 2015). Infrastructure is often reduced to an input variable or background condition, typically measured through investment levels or building standards without taking into account students' lived experiences of their physical environments (Maxwell, 2016; Belmonte, et al., 2019). This approach limits the explanatory power of current models and obscures the role infrastructure plays as a co-constitutive element of the educational system.

In response to these limitations, this study adopts a complexity-informed theoretical perspective. Complexity theory conceptualises schools as dynamic systems in which outcomes emerge from the interaction of multiple, interdependent elements (Mason, 2016). From this view, infrastructure cannot be treated as a stable input with predictable effects; instead, it operates as a relational factor within broader networks of influence, interacting with school culture, leadership, pedagogy, and students' social experiences. Complexity theory, therefore, provides an appropriate foundation for interpreting the unpredictable, non-linear, and context-dependent patterns through which infrastructure may shape outcomes (Wrigley, 2019).

The study draws on and operationalises the School Climate framework developed by Wang and Degol (2016). This framework conceptualises School Climate as the overall quality of school life, encompassing the four pillars of Safety, Interpersonal Relationships, Teaching and Learning, and Institutional Environment. While existing applications of the framework focus predominantly on social and pedagogical elements, this study extends it to include infrastructure as a core domain, which is consistent with the proposition that physical environments shape and are shaped by school culture (Uline & Tschannen-Moran, 2008; Rudasill, et al., 2018).

Methodologically, the study employs machine learning techniques, cluster analysis and gradient-boosted decision trees to explore non-linear and context-specific associations between infrastructure and educational outcomes. These techniques are well suited to identify emergent patterns within complex systems, as they do not require assumptions of variable independence or linearity (Zawacki-Richter, et al., 2019; Tin, et al., 2024). Rather than seeking deterministic causal chains, the study uses machine learning to surface meaningful patterns of association that reflect the adaptive and contingent nature of educational environments (van Vondel, et al., 2016; Yang, 2024).

These theoretical and methodological commitments lead directly to the central research question:

RQ: *What is the relationship between school infrastructure and educational outcomes?*

This question synthesises the study's interest with the relational nature of infrastructure; the emergent character of educational outcomes; and the need for analytical tools capable of capturing complexity. It also frames two subsidiary research questions that clarify the distinct contributions the study seeks to make.

The first subsidiary question addresses the need to reposition infrastructure within students' lived experiences rather than treating it as an external or background factor:

SRQ 1: *How does situating infrastructure within the broader context of student experience inform the understanding of its role in shaping educational outcomes?*

This question enables an exploration of the relational and perceptual dimensions of infrastructure within school life, thereby contributing to a more holistic account of its educational significance.

The second subsidiary question focuses on the interpretive value of adopting a complexity-informed perspective:

SRQ 2: *To what extent does a complexity-informed approach help to assess the contribution of infrastructure to educational outcomes?*

The study evaluates whether complexity theory offers an appropriate theoretical and practical lens for examining the multifaceted and dynamic relationships between infrastructure and educational processes.

Together, these research questions define the study's conceptual and analytical architecture. They emerge directly from the literature, reflect the study's epistemological foundations, and inform both the design and analysis set out in the following chapters.

Chapter 3. Method

This study adopts a pragmatist philosophical orientation, which was selected for its capacity to accommodate methodological pluralism and support context-sensitive, outcome-focused inquiry. Pragmatism offers a flexible epistemological stance enabling researchers to draw on quantitative and qualitative elements, unencumbered by the philosophical rigidity of positivist or constructivist paradigms (Shook, 2023).

This study investigates the relationship between school infrastructure and educational outcomes (RQ1), with a particular focus on students' lived experiences of infrastructure as situated within the broader ecology of school life. By centring student perception, the research foregrounds the experiential dimension of infrastructure (RQ1.1) and acknowledges the non-linear, interdependent relationships characteristic of complex school environments (RQ1.2). To operationalise these constructs, a modified and expanded School Climate survey instrument is employed, which incorporates a five-point Likert scale to collect data on both student experiences of infrastructure and associated educational outcomes.

Given the limitations of traditional statistical techniques in modelling complex, non-linear relationships, this study adopts machine learning methods, specifically cluster analysis and GBTs, to analyse the student experience and outcome data. While typically associated with positivist paradigms, the application of machine learning techniques in this study is informed by a pragmatist commitment to methodological flexibility and real-world relevance. GBTs are well-suited to addressing the research question, facilitating the identification of complex patterns between students' perceived experience with infrastructure and educational outcomes.

The research methodology unfolds across three interrelated stages: development, design, and execution. The *development* stage establishes the conceptual and philosophical foundations of the study, ensuring coherence between the research question, context, and methodological choices. The *design* stage details the construction of the survey instrument, sampling procedures, and data preparation and exploration strategies. Finally, the *execution* stage provides a comprehensive account of the analytical techniques employed and the rationale behind the selection of specific tools for prediction, clustering, and interpretation. A visual representation of these three stages is provided in Figure 3.1.

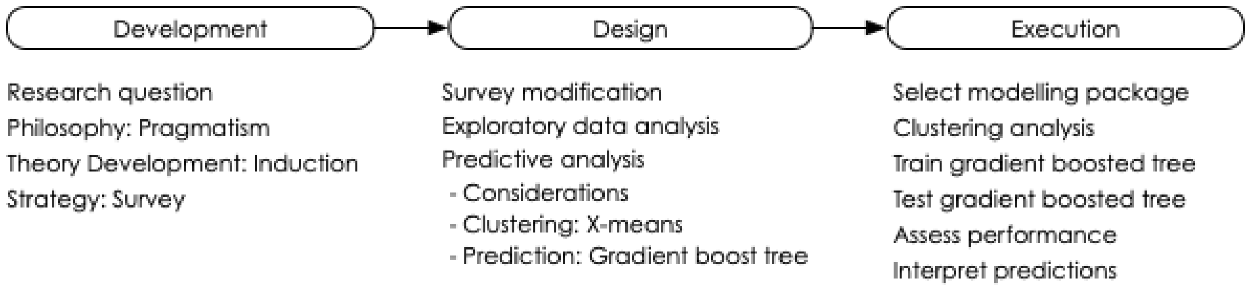

Figure 3.1. Overview of the development, design, and execution stages of the method

3.1 Development

This chapter begins by articulating the philosophical foundations of the research within the pragmatist paradigm. It then presents a detailed account of the study's methodological design, including data collection instruments and analytical procedures. The chapter also discusses ethical considerations, limitations, and the implications of the chosen approach, reinforcing the alignment between methodology and the overarching research objectives.

3.1.1 Research philosophy: Pragmatism

The study is grounded in the pragmatist paradigm, characterised by its practical orientation, methodological pluralism, and rejection of rigid philosophical binaries (Tashakkori & Teddlie, 2010). The emphasis of pragmatism on "what works" enables researchers to effectively integrate methods from diverse paradigms to address complex research questions (Creswell, 2009).

Pragmatism emerged in the late 19th century through the work of philosophers such as Charles Sanders Peirce, William James, and John Dewey (Gillespie, et al., 2024), who advocated for a practical, outcomes-focused approach to knowledge generation. This philosophy rejects the extremes of positivism and interpretivism, instead advocating for a middle ground that values both empirical observation and contextual interpretation (Shook, 2023). Contemporary scholars such as Biesta (2012) and Goldkuhl (2012) have further developed pragmatism as a research philosophy, particularly in studying complex and dynamic systems.

3.1.1.1 Ontological and epistemological assumptions

Pragmatism adopts a flexible ontological stance that resists strict dichotomies between objectivism and subjectivism (Allemang, et al., 2022). Pragmatism views reality as external and constructed, shaped by interactions between individuals, systems, and environments (Morgan, 2014). This perspective aligns with the conceptualisation of schools as complex systems, where outcomes emerge from the interplay of structural, relational, and contextual factors (Bloom, 2019). Schools exhibit complex attributes, including feedback loops, non-linearity, and emergent behaviours (Duit, et al., 2010). Pragmatism accommodates this complexity by recognising that reality is not static but evolving in response to ongoing interactions (Morgan, 2014).

Epistemologically, pragmatism prioritises generating practical, actionable, and context-sensitive knowledge (Saunders, et al., 2016). It resists the positivist quest for universal, generalisable truths, instead valuing knowledge for its ability to address specific challenges and inform decision-making (Baima & Paytas, 2020). Pragmatism supports integrating quantitative and qualitative methods by recognising that different tools can yield complementary insights (Johnson & Onwuegbuzie, 2004). This study employs quantitative methods not to assert deterministic causality but to identify and generate insights relevant to specific school contexts.

3.1.1.2 Relevance to educational research

Pragmatism is particularly suited to studying educational systems characterised by complexity and variability (Koopmans, 2020). Traditional positivist approaches often struggle to capture the nuanced interplay of the factors shaping educational outcomes (Biesta, 2010; Foster, 2023). Importantly, identified relationships linking inputs and outcomes, reliant on reductionist assumptions (Geyer, 2012), typically have little value in shaping successful policy (Goldspink, 2007; Fullan, 2019). Pragmatism, by contrast, supports a more holistic approach and enables researchers to explore relationships, interactions, and emergent phenomena without being constrained by reductionist assumptions (Ansell & Geyer, 2017; Shook, 2023).

In considering the evolution of educational research, Lyu (2024) notes the growing shift away from either purely positivist or interpretivist paradigms to enable combined qualitative and quantitative approaches “ensuring a comprehensive understanding of complex educational phenomena” (p. 11). The focus of pragmatism on practical outcomes aligns with this study’s objectives of contributing knowledge with the potential to inform approaches to school improvement and policy decision-making.

3.1.1.3 The contribution of positivism

Although the study adopts a pragmatist philosophy, elements of the methodology draw on positivist traditions, particularly in their reliance on numerical data and structured analysis. Positivism, originating from the work of Auguste Comte and further developed in the 20th century, posits that reality is objective, measurable, and governed by universal laws (Bryman & Bell, 2015). This paradigm emphasises empirical observation and statistical analysis as the primary means of generating knowledge (Raines, 2013).

Positivist approaches are prevalent in educational research, particularly in studies examining relationships between input variables, including infrastructure and student outcomes (Bromley, et al., 2023). Techniques such as regression analysis, structural equation modelling, and effect size calculations exemplify the positivist emphasis on establishing causal relationships (Simpson, 2018). These methods enable researchers to identify patterns, generalise findings across contexts, and provide insights into the factors shaping educational performance.

Despite its prevalence and ongoing contributions to educational policy discourse, positivism faces significant limitations when applied to complex systems such as schools (Loughland & Thompson, 2016; Lyu, 2024). Scholars such as Sterman (2000) and Goldspink (2007) argue that positivist methods generally oversimplify the intricate dynamics of educational systems and treat schools as mechanistic entities rather than adaptive, emergent systems. Policies derived from positivist research frequently fail to achieve their intended outcomes because they overlook contextual variability (Goldkuhl, 2012; Waslander, et al., 2020) and systemic feedback loops (Biesta, 2010; Fullan, 2019). Further, Geyer (2012) argues that the failure of positivism in the context of educational policy has led to an increasing level of audit and compliance, which destabilises and worsens educational outcomes.

In this study, the use of survey data, cluster analysis, and predictive modelling reflect elements of positivism. The survey data is numeric, reflecting the responses to the five-point Likert scales. Predictive analytics techniques are used to quantitatively analyse this data. However, the data and analysis are employed within a broader pragmatist framework that prioritises their utility for addressing the research questions. Specifically, the intent in analysing the relationships between inputs and outcomes is not to generate causal relationships informing the development of improvement policy. Instead, the intent is to understand the nature of the relationships between the inputs and outcomes, including whether they are, as some researchers posit, complex and school specific. The study leverages the strengths of the positivist approach in analysing structured data by situating quantitative tools within a context-sensitive approach.

3.1.1.4 Methodological alignment with pragmatism and complexity theory

The emphasis of pragmatism on outcomes, flexibility, and methodological pluralism (Morgan, 2014) underpins the study's methodological choices (Shook, 2023). Quantitative tools are employed not as ends but as a means of generating insights into the School Climate attributes that most influence educational outcomes.

The study uses a survey instrument to collect data on School Climate and related factors, with respondents providing numeric responses on a five-point Likert scale, which facilitates data standardisation and supports quantitative analysis. The data used in this study, specifically the student survey responses, is numerical. The survey design reflects the focus of pragmatism on methodological appropriateness, balancing the need for comparability with the flexibility to capture diverse perspectives.

Cluster analysis groups schools into outcome classes based on shared characteristics. This technique identifies latent patterns within the data, revealing relationships that may not be apparent through traditional linear methods. Cluster analysis aligns with the pragmatist emphasis on exploring complex interactions while avoiding deterministic assumptions.

Machine learning classification algorithms, specifically GBTs, predict the outcome classes derived from the cluster analysis. This machine learning technique is particularly suited to handling non-linear relationships and interactions, making it ideal for studying complex systems such as schools. GBTs provide interpretable models that highlight key variables and their interactions, which enables the generation of actionable insights for decision-making. Unlike traditional regression methods, GBTs capture the multifaceted nature of educational systems without imposing linear assumptions.

Schools are conceptualised as complex adaptive systems characterised by non-linearity, emergent behaviours, and context dependence (Duit, et al., 2010; Snyder, 2013). Complexity theory highlights the importance of understanding relationships, interactions, and feedback loops within systems, challenging traditional reductionist approaches (Koopmans, 2020).

Pragmatism provides a philosophical framework accommodating complexity (Ansell & Geyer, 2017; Koopmans, 2020), enabling researchers to integrate diverse methods and perspectives (Morrison, 2010). By combining survey data, cluster analysis, and machine learning, the study captures the multidimensionality of School Climate while generating insights tailored to specific contexts. This approach reflects the emphasis of pragmatism on actionable knowledge (Allemang, et al., 2022).

3.1.2 Approach to theory: Induction

This study adopts an inductive approach to developing and exploring theory. This section explains the main approaches used in developing the theory and the choice of an inductive approach for this research.

Inductive, deductive, and abductive approaches are the three methods used to develop theories explored through research (Hurley, 2011). Each approach offers a unique perspective and procedures for generating and validating theories. The attributes of these approaches play a critical role in shaping the design and conduct of studies (Locke & Latham, 2020). Understanding the differences between these approaches is essential for appropriate theory development.

The inductive approach involves moving from specific observations or empirical data towards broader generalisations and theories. In this method, researchers start with a collection of data (Shrestha, et al., 2021), which can be obtained through quantitative means or qualitative methods such as interviews, observations, or content analysis (Locke, 2015). This data is then analysed with the intent of identifying patterns, themes, and relationships. Researchers aim to develop a theory that can explain the observed phenomena by systematically categorising and comparing these patterns. While traditionally associated with qualitative analysis, Faems (2020) argues that with improvements in machine learning it is increasingly possible to place quantitative data at the core of inductive theorising.

Inductive reasoning is often more exploratory and open-ended, as researchers do not start with a predefined theory or hypothesis; instead, they allow the theory to emerge from the data itself. While this approach allows for the discovery of novel insights and unexpected connections, it may also pose challenges in ensuring the rigour and general applicability of the resulting theory, as it relies significantly on the quality and representativeness of the data (Hurley, 2011).

This study adopts an exploratory approach, employing machine learning algorithms to analyse School Climate data to deepen the understanding of how infrastructure influences educational outcomes across diverse school environments. In assessing the role of machine learning in theory development, Shrestha et al. (2021) argue that such methods remain undervalued and underexplored, despite their potential to support inductive theory-building through data-driven analysis.

This study adopted an inductive approach to align with its exploratory focus and to allow for methodological flexibility as insights emerged during the data analysis process. While this approach limits the immediate development or testing of formal theory, since theory confirmation requires subsequent deductive research (Creswell, 2009), such a limitation is acceptable in this exploratory context. The study may be understood as a preparatory stage, laying the groundwork for future research aimed at establishing and testing specific theoretical frameworks.

3.1.3 Strategy: Survey

To support its exploratory and inductive approach, this study uses a survey strategy designed to enable systematic, large-scale data collection. As defined in methodological literature, a survey strategy involves systematically collecting data from a defined population, often through questionnaires, to gather information about behaviours, attitudes, or perceptions (Creswell, 2009; Groves et al., 2014). In the context of educational research, survey strategies are particularly effective when the aim is to investigate patterns across diverse school environments using a structured and repeatable instrument (British Educational Research Association, 2017). Both pragmatic and methodological considerations underpin the decision to employ a survey.

First, the research seeks to understand how students experience school infrastructure (an element of School Climate) and how these perceptions relate to educational outcomes. Surveys are commonly used in School Climate research (Thapa, et al., 2013; Wang & Degol, 2016), including when trying to understand the effect of infrastructure on School Climate (Uline & Tschannen-Moran, 2008). Surveys are suitable for capturing such perceptions in a standardised format, allowing the researcher to gather comparable data from a large population (Groves, et al., 2014). As Daniel and Harland (2017) note, surveys enable quantitative researchers to examine how respondents have "experienced, interpreted or understood" aspects of education (p. 45), which makes them a valuable tool in research engaging with subjective constructs through measurable variables.

Second, the survey method aligns with the study's machine learning techniques, which require large, structured datasets to identify patterns and support inductive insight generation (Larose & Larose, 2015). While the study does not aim to test an existing theory, it does seek to uncover relationships within the data that may inform future theoretical development. Shrestha et al. (2021) observe that machine learning remains underutilised in theory-building despite its potential to support inductive analysis in data-rich contexts. By adopting a survey strategy, this research generates the structured, high-volume dataset required to apply such techniques.

Finally, surveys provide practical advantages in understanding the complexities of educational systems. They allow for consistent data collection across varied contexts while accommodating diverse student experiences (Yorke, 2009). The selection of a School Climate survey is intended to support the investigation of systemic responses to structural factors, including infrastructure, across different school environments.

3.2 Design

This section describes the process whereby an existing School Climate survey formed the foundation for the data collection process. This survey, modified to align with the chosen School Climate framework, was then administered and the responses made available for analysis. This process is illustrated in Figure 3.2, followed by a detailed explanation.

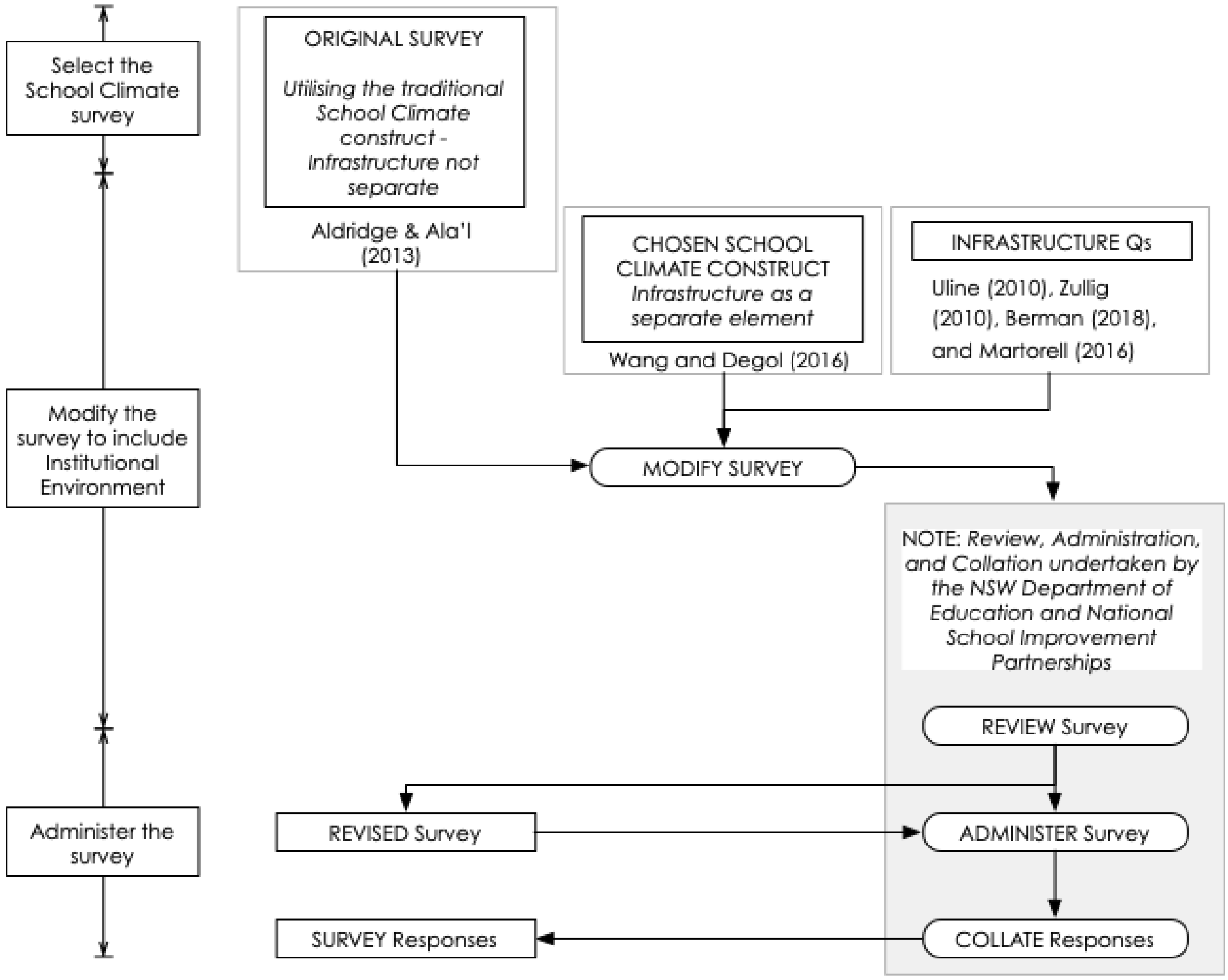

Figure 3.2. Overview of the selection, modification, and administration of the School Climate survey

In 2022, the NSW Department of Education, through School Infrastructure New South Wales (SINSW), initiated the administration of School Climate surveys to inform the planning of upcoming infrastructure upgrades. SINSW deemed the understanding of School Climate to be a fundamental initial step in engaging with schools earmarked for potential redevelopment.

An existing School Climate survey was expanded to capture the perceptions of students, staff, and other stakeholders within the school community regarding the influence of infrastructure on their school experience. The insights obtained from this initiative were intended to support decision-making processes by providing empirical evidence on how existing infrastructure conditions impact the lived experiences of school communities.

Although the primary objective of the survey was to assist departmental planning, the data it produced closely resonated with the focus of this research, that is, understanding the contribution of infrastructure to educational outcomes when viewed through the lens of complexity. Consequently, this study leverages the survey data to investigate the relationship between school infrastructure and educational outcomes.

It is essential to acknowledge that the researcher played a role in modifying the survey instrument. The modifications were implemented to meet the strategic and operational requirements of the NSW Department of Education. They were developed in consultation with the researcher responsible for the original survey and were aligned with the department's objectives relating to infrastructure planning and school engagement. The modifications focused on relevance, clarity, and usability within the department's planning processes. This section describes the development of the revised survey instrument.

3.2.1 Development of the survey instrument

Numerous School Climate surveys exist and form the basis for ongoing research in this field. However, none of the School Climate surveys validated in Australia or available and approved for administration in NSW government-funded schools contained questions covering infrastructure. As such, an existing School Climate survey was extended by including appropriate questions to address the Institutional Environment. As noted in 2.4.2, the definition of Institutional Environment for this study is restricted to infrastructure, whereas the definition proposed by Wang and Degol (2016) is broader and includes the school's organisational structure and access to resources.

The School Climate survey developed by Aldridge and Ala'I (2013) is the source of the non-infrastructure questions for this study. A key factor in the decision to use this survey was the need for the survey to be approved for administration by the NSW Department of Education. The selected survey is one of a suite of tools developed and validated in Australia in work led by Aldridge and collaborators (Aldridge & Galos, 2018; Aldridge & McChesney, 2021; Aldridge & Ala'I, 2013). In addition to secondary schools, these surveys cover parents and caregivers (Aldridge & McChesney, 2021), primary students (Aldridge & Galos, 2018), and teachers (Aldridge & Fraser, 2018). The School Climate framework underpinning the work of Aldridge uses the traditional constructs of Safety, Interpersonal Relationships, and Teaching and Learning. Infrastructure experience, however, is not part of this source survey.

In their review of the literature covering School Climate, Wang and Degol (2016) provide early documented consideration of infrastructure as a standalone element in the School Climate construct. Wang and Degol categorise infrastructure as the Institutional Environment, which is defined as "organisational or structural features of the school environment" (2016, p. 317). Wang and Degol noted the importance of Institutional Environment in reviewing the literature; however, they did not examine specific questions related to infrastructure in their review.

Existing studies that address the contribution of infrastructure to educational outcomes informed the scope of the infrastructure experience required in the Institutional Environment section of the survey instrument. The sources detailed in Table 3.1 contributed to the questions used in the Institutional Environment section of the survey. However, the questions were modified specifically for this research undertaking to ensure a consistent style, tone, and level throughout the survey. Individual questions developed for the Institutional Environment section of the survey were divided into five sub-constructs: ambient environment; ergonomics; indoor design; outdoor design; and condition, maintenance, and upkeep. This categorisation provides a more coherent survey experience for the respondents and adds clarity and context to later reporting and analysis.

When developing questions for the Institutional Environment section of the survey, sources that examined infrastructure within the context of School Climate or user experience were prioritised. The work of Zullig et al. (2010) focuses on using historical literature to create a valid and reliable student-reported School Climate instrument. Among the five defined domains, infrastructure is captured within School Facilities, which encompasses the sub-domains of temperature, noise levels, physical condition, classroom layout, and decorative elements.

As shown in Table 3.1, the work of Zullig et al. (2010) contributed to three questions. Questions relating to arrangement and decoration were omitted.

Table 3.1. Sources for the Institutional Environment survey questions

Sub-construct	Statements	Source					
		Zullig et al (2010)	Uline and Tschannen-Moran (2008)	Karippanon (2018)	Thapa et al (2013)	Department of Education *Tell Them From Me survey*	New summary question
Condition, maintenance, and upkeep	Building exterior well maintained.		Y			Y	
	Classroom furniture and flooring are well maintained.		Y				
	Grounds well maintained.	Y				Y	
	Inside of school buildings well maintained.	Y				Y	
	School is kept clean and tidy.	Y	Y			Y	
Ambient environment	Temperature is about right.		Y				
	There is a pleasant smell or no smell.		Y				
	The amount of light is about right.		Y				
	Not disturbed by noise in study time.		Y				
	I can see what is displayed without difficulty.			Y			
Ergonomics	The chairs and other seating are comfortable.			Y			
Indoor design	Design of learning spaces meets my learning needs.		Y				
	The classrooms and other learning spaces provide a welcoming atmosphere.		Y				
	Classrooms and learning spaces provide enough space.		Y				
	Classrooms and learning spaces adequate for different subjects.		Y				
	I am generally satisfied with the classrooms and other learning spaces at the school.		Y				Y
Outdoor design	There is enough equipment and courts (e.g. tennis courts, basketball hoops) for all who want to use it.				Y	Y	
	There are sufficient areas for students who want to be active and noisy.				Y	Y	
	There are sufficient spaces for relaxing, reading, or quiet reflection.			Y			
	There are sufficient spaces where students can socialise (e.g. places to sit together).			Y			
	Enough shelter for protection from the wind and rain.					Y	
	There is enough shelter for protection from the sun on hot days.					Y	
	Outdoor spaces offer a sufficient selection of activities.						Y

In examining the classroom arrangement, Rogers (2020) found that what constitutes an effective classroom arrangement can vary among subjects, teaching styles, and pedagogy. The NSW Department of Education determined that collecting data for each subject-specific classroom was not feasible. Likewise, decoration, which includes colours and subject-specific material (Barrett, et al., 2013), was omitted from this study following guidance from the NSW Department of Education.

The work of Uline and Tschannen-Moran (2008) contributes significantly to the survey used in this research. Uline and Tschannen-Moran do not consider infrastructure to be a component of School Climate; instead, they find that infrastructure quality can moderate School Climate. Their work is important in framing infrastructure quality as perceived by the user and, as such, in a way compatible with the School Climate. However, it is important to note that Uline and Tschannen-Moran's study surveyed teachers, not students. Their survey questions required modification to account for the differences between teachers and students, and to ensure consistency with the source School Climate survey. While the original questions effectively captured teacher perceptions of infrastructure quality, the negative framing and generality of statements such as "The facilities here are lacking in regular maintenance" (p. 63) are not optimal for student responders. Specifically, negative framing (e.g. "lacking") has been shown to increase cognitive load and lead to inconsistent or biased responses, as respondents mentally construct a positive scenario first before negating it (Clark & Chase, 1972, p. 473). Additionally, the broad nature of such statements (e.g. "here") does not account for the specific environments that students interact with daily, such as school grounds, indoor learning spaces, and building exteriors.

To accurately reflect the students' experiences, the questions were reframed positively and made explicit for each environment. Positive statements, such as "The school grounds are well maintained", reduce cognitive complexity and enable respondents to provide clearer and more reliable feedback (Just & Carpenter, 1971). Furthermore, dividing the survey into specific domains ensures that the collected data reflects students' varied interactions with different aspects of infrastructure (Paivio, 1990, p. 72). These modifications preserve the intent of the original work by Uline and Tschannen-Moran (2008) while tailoring it to suit the survey respondents and the source School Climate survey.

Kariippanon et al. (2018) focused on the evolution of classroom spaces to better suit the learning needs of 21st-century school students without being subject-specific. Two questions were derived from this work, covering the students' ability to see the presented information and the comfort achieved when seated in lessons.

Administered by the NSW Department of Education, the *Tell Them From Me* survey (NSW Department of Education, 2023) elicits views from students, teachers, and parents. In 2019, the survey introduced infrastructure questions as an initial effort to complement objective assessments of infrastructure condition and functionality with students' views. Infrastructure questions from the *Tell Them From Me* survey remain unchanged, as they were consistent with questions asked in other source surveys.

As shown in Table 3.1, two questions, denoted as summary questions, are not sourced from existing surveys. Consistent with the primary source survey, these summary questions are an internal consistency check.

Before administration of the survey, SINSW reviewed the survey instrument in its entirety. The review ensured that the language used in the survey instrument was consistent with that used in NSW schools and that the survey participants would easily understand the questions. Reviewers included teachers, NSW Department of Education legal staff, and senior bureaucrats from within the department. Numerous small wording changes aligned the survey with departmental language, including replacing the term bathroom with toilet facilities. Reviewers also recommended including additional details for some questions. For example, instead of asking if "There are sufficient spaces where students can socialise", the administered version reads "There are sufficient spaces where students can socialise (e.g., places to sit together)". Survey responses generated by the reviewers were not retained or included in the responses used in this research.

The survey as approved for administration by the NSW Department of Education is included in Appendix 1. Two separate approvals were required before the survey could be administered and the responses used in this research. First, the UTS Human Research Ethics Application (ETH22-7523) was approved and assessed as Low Risk and approved by the Ethics Secretariat in February 2023. The NSW Department of Education administered the School Climate survey and managed respondent information, instructions, and consent. As such, these administration elements, detailed in section 3.2.3, did not form part of the ethics submission. Second, the NSW Department of Education provided formal approval for the survey data to be used in the research on 15 September 2022 (NSW Department of Education, 2022), this letter is included at Appendix 2.

3.2.2 Survey: Selecting the schools

When this research project entered the data collection stage, School Climate survey results were available for 29 schools throughout NSW. These schools had been selected by the NSW Education Department either because they were under consideration for infrastructure upgrades or because they were of particular interest to the Deputy Secretary of Public Schools. In the latter instance, the selection criteria included relatively poor academic performance, unfavourable media attention, or a rising incidence of anti-social behaviour.

The department sought to develop a deeper understanding of how students, teachers, parents, and non-teaching staff experienced the impacts of infrastructure upgrades. Unlike the annual *Tell Them From Me* survey (The Learning Bar, 2024), which originated as a School Climate survey, this survey aimed to analyse data at the school level rather than at the system level. However, following receipt of the survey data, the focus shifted to the ten secondary schools. Two primary considerations drove this decision. Firstly, the constructs of the student survey differ between secondary and primary schools, with the secondary school survey specifically designed to capture students' current and preferred school experiences in three of the four School Climate pillars. This difference between the current and preferred experience, known as climate misfit, is an important indicator of the nature and strength of the students' perceived experience. Secondly, the anticipated number of responses influenced the decision to focus on secondary students. Specifically, public (primary) schools are typically smaller than secondary schools and there are three separate survey instruments for public schools, each covering students in a different two-year window. As such, the number of responses in each public school cohort (Lower, Middle, and Upper) is likely to be insufficient for the intended numerical analysis. Specifically, non-linear techniques, which are essential for analysing the complex and dynamic nature of schools, require a substantial amount of data. Parents, caregivers, teachers, and non-teaching staff are excluded for this reason.

3.2.3 Survey: Administration

Departmental staff briefed principals and deputy principals from each of the selected schools. The purpose of the briefings was to explain the survey as the first step in better understanding the views of students, teachers, non-teaching staff, parents, and caregivers. A dedicated team from the department liaised with each selected school, setting aside dedicated time on specified days. The departmental team then visited each school, providing electronic tablets for all students. Students completed the surveys on a bespoke platform owned and operated by National School Improvement Partnerships (NSIP) (2023). NSIP is a commercial entity launched by Curtin University to enable more expansive use of the university's School Climate surveys at more commercially acceptable rates. NSIP then processed the completed surveys.

Surveys were offered to all school users; however, participation was voluntary with permission managed by the NSW Department of Education. NSIP and departmental staff briefed all potential participants and were available to answer questions and assist participants with any technical issues. Following administration of the survey, this research process diverged from the ongoing departmental work. Specifically, NSIP produced school-specific reports that were briefed to the department and the principals, forming the basis of coordinated whole-of-school improvement efforts by the department and the individual schools. The department was not provided with the survey data in any form other than the final reports for each school. This data, following written agreement from the department and in accordance with the Research Ethics Submission, was provided to the researcher in CSV format. Individual anonymous responses were provided for all participants. The data was provided by NSIP before any cleansing or quality checking, except for the automated checks at the point of administration.

3.2.4 Survey: Respondent participation

Table 3.2 details the number of responses, response rate, and relevant contextual information for the ten secondary schools participating in the School Climate survey. A total of 5,291 students from ten schools responded to the survey, which is a response rate of 64%. It is important to note that the contextual information, including that describing the student population, is sourced from the *My School* website (Australian Curriculum Assessment and Reporting Authority, 2020)administered by the Australian Curriculum, Assessment and Reporting Authority (ACARA), an independent statutory authority that provides the *My School* website to support "national transparency and accountability of Australia's school educational system through publication of nationally consistent school-level data" (Australian Curriculum Assessment and Reporting Authority, 2020).

Table 3.2. Respondent participation for the ten secondary schools participating in the School Climate survey

School	Number of responses	School enrolment	Response rate	Percentage of students enrolled who identify an Indigenous	Percentage of students with a language background other than English	Index of Socio-Educational Advantage (ICSEA)	Attendance overall rate	Students attending more than 90% of the time
A	411	685	60%	18%	6%	26	79%	37%
B	361	636	57%	41%	6%	4	68%	27%
C	783	1223	64%	2%	70%	76	88%	57%
D	611	878	70%	0%	94%	21	80%	27%
E	193	371	52%	20%	41%	6	74%	30%
F	527	780	68%	5%	8%	62	84%	36%
G	448	655	68%	20%	9%	16	74%	27%
H	618	1036	60%	12%	7%	40	84%	47%
I	582	880	66%	3%	80%	25	84%	43%
J	757	1139	66%	8%	7%	41	84%	41%
Mean	5,291	8,283	64%	13%	33%	32%	80%	37%

There are two reasons for using external, government data to describe the surveyed schools and their student populations. First, the NSW Department of Education raised concerns about the potential identifiability of students in smaller schools when survey responses include demographic information. This issue directly relates to privacy and child protection laws in NSW, notably the *Privacy and Personal Information Protection Act 1998* (NSW) (PPIPA) (NSW Government, 1998a) and the *Children and Young Persons (Care and Protection) Act 1998* (NSW) (NSW Government, 1998b). These laws mandate stringent measures to safeguard student confidentiality and welfare. Due to limited enrolment, small schools inherently increase the risk of identifying individual students based on demographic combinations. That is, a specific gender, age, or cultural background might correspond to only one or two students in the school, which makes identification possible even without personal identifiers. This risk of re-identification undermines privacy safeguards and can lead to breaches of confidentiality, which the PPIPA aims to prevent. Secondly, in some cases, the survey enabled free text responses and omissions in areas covering demographic and other contextual data. As such, the publicly available, government endorsed data was used to ensure data quality.

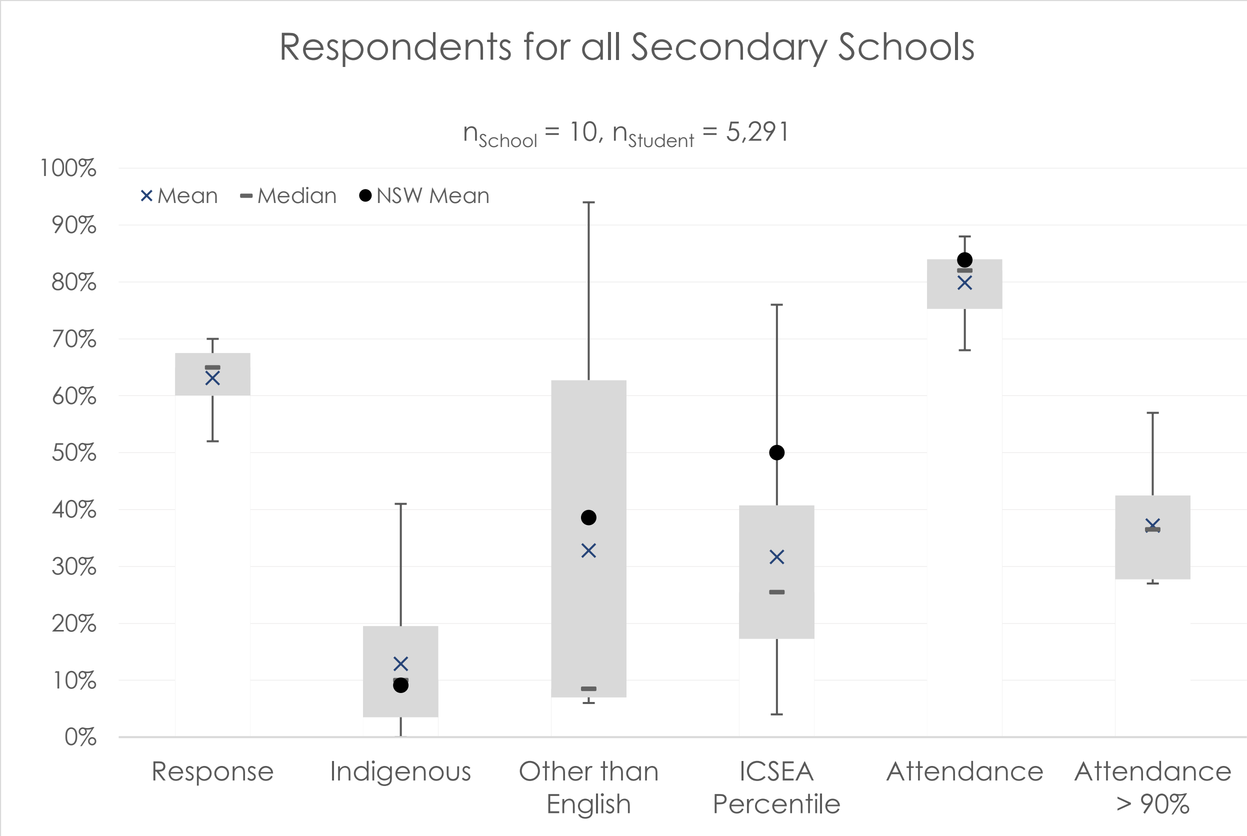

Figure 3.3 presents the descriptive statistics covering the response rates and school contextual data. The contextual data enables an assessment of how representative the sample of schools is compared to all schools in NSW, except for that data detailing the percentage of students attending school for nine out of every ten days (*Response* and *Attendance > 90%*). In regards to Attendance > 90%, no departmental (or other official) data enabling whole-of-system comparison is available.

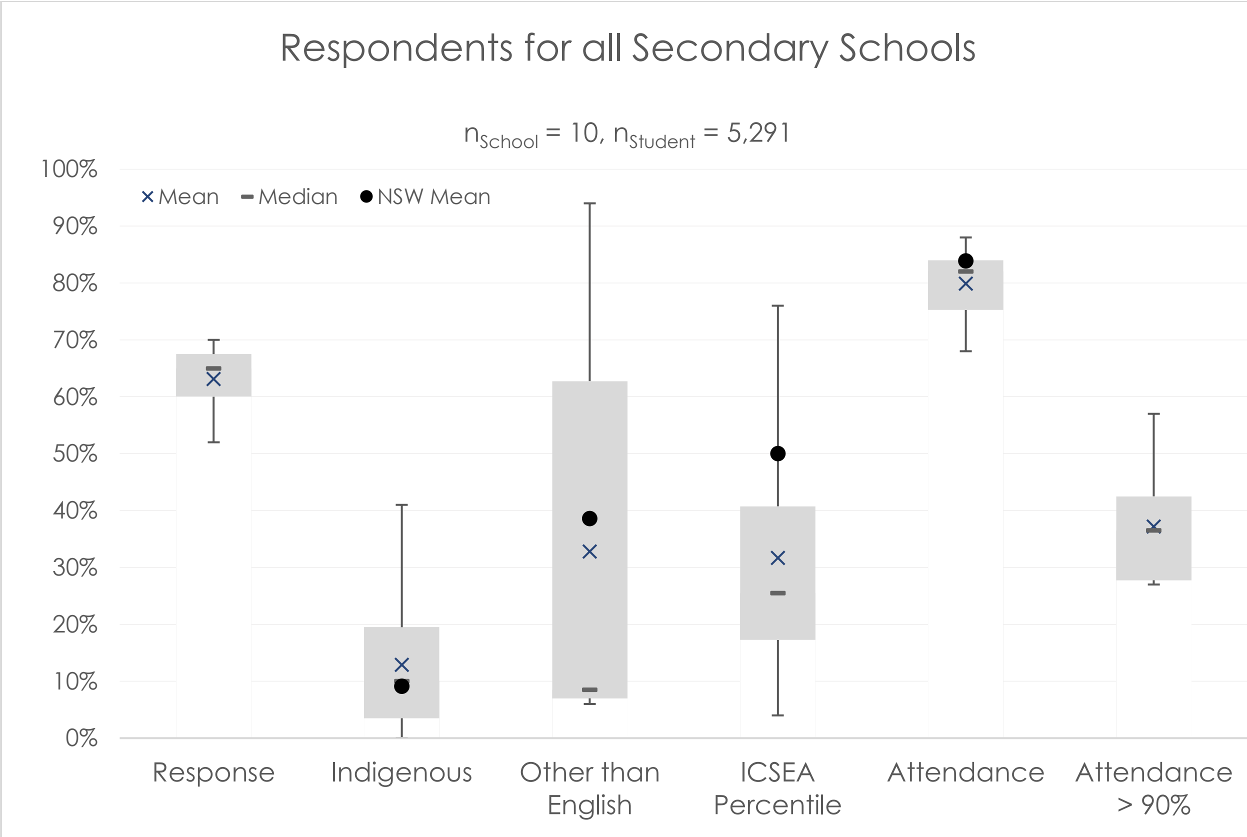

Figure 3.3. Descriptive statistics covering the response rates and school contextual data compared to the NSW mean data

The surveyed schools' mean percentage of Indigenous student enrolments is above the NSW mean of 9.3%. The surveyed school cohort has a lower than average percentage of students with a background other than English compared to the NSW mean of 38.6% (Centre for Education Statistics and Evaluation, 2024). However, the intra-cohort range for backgrounds other than English is significant, with a low of 6% (schools A and B) and a high of 94% (School D).

The Index of Community Socio-Educational Advantage (ICSEA) is a socio-educational advantage scale calculated by ACARA (2020) using factors such as parental education and occupation, and school characteristics such as location and the students' socio-economic background. The ICSEA percentile (ranging from 1 to 100) measures relative advantage. For example, a school reporting an ICSEA percentile of 40 is, compared to all schools, *more advantaged* than 40% and *less advantaged* than 60% of schools (2020). Although the surveyed school group is diverse, it does not constitute a representative sample of NSW schools. Notably, the cohort's mean ICSEA percentile is 32%, suggesting a relatively disadvantaged profile compared to the NSW average. However, the purpose of this research is to explore predictive relationships between School Climate and educational outcomes (wellbeing and behaviour) both within and between schools. Consequently, a statistically representative sample of the NSW secondary school population is not necessary for this analysis.

3.2.5 Exploratory data analysis

The first analytical step in this process is exploratory data analysis (EDA). The intent of EDA is to delve into the data in order to understand the attributes and their relationships, and the relationships between predictors and the predicted targets (Larose & Larose, 2015).

The survey data provided was limited to the School Climate construct. Publicly available contextual information from the *My School* website and the Australian Bureau of Statistics (2017) provided additional contextual information. The additional data covered enrolments, attendance, and socio-educational and socioeconomic status.

Considering the source of the data and surveys, most attributes are categorical rather than numerical. EDA of categorical variables from the surveys identifies the differences and similarities in the responses provided by different user types across the schools. As the surveys were administered to secondary students, they contained a section where respondents detailed their current and preferred experience, specific to different elements. This is known as the School Climate misfit (Aldridge, et al., 2024).

Two essential components of EDA are distribution analysis and comparison analysis, the purpose of which is to examine the initial information provided by the respondents and to identify relationships of interest in the second stage, the predictive analysis.

Some aspects of the What's Happening In This School (WHITS) survey involve highly correlated responses, which is an essential part of the initial validation process (Aldridge & Ala'I, 2013). The response analysis aims to understand how much the data can support predictive analysis. Specifically, when designing the predictive analysis, the degree of correlation between inputs, between outputs, and between inputs and outputs is essential knowledge.

The exploratory analysis also provides information critical to the design of predictive analysis. Different predictive techniques require different amounts of data. As such, the number of responses in each cohort is a significant factor in determining the level at which predictive models operate. Modelling at the individual school level is ideal. However, if insufficient data is available, options such as rejecting schools or combining cohorts must be considered, explained, and justified. Finally, the response analysis details those attributes where the responses contain errors or missing values. That is, a crucial function of EDA is assessing the data quality.

3.2.6 Student outcomes: Clustering to produce outcomes categories

The administered survey is designed to elicit both student experience and reported outcomes. There are two broad outcome constructs: wellbeing and behaviour. These two constructs are then divided into more detailed sub-constructs. Five sub-constructs describe wellbeing: self-efficacy, vigour, resilience, learning goal orientation, and social harm. Two sub-constructs describe behaviour: disruptive behaviour and risky behaviour. Students report their outcomes against these seven sub-constructs by responding to 28 statements, as detailed in Table 3.3. The output data consists of responses to questions answered using a five-point Likert scale.

Table 3.3. The 28 outcome statements covering the constructs wellbeing and behaviour

Construct	**Sub-construct**	**Outcome statement**
Wellbeing	Self-efficacy	At school, even if the work is hard, I can learn it.
		I feel that I will achieve a good result at school.
		At school, I can do difficult work if I try.
		I understand what is taught at school.
	Vigour	I am determined to achieve my goals.
		I bounce back after difficult times.
		I can achieve goals despite barriers.
		I come through difficult times with little trouble.
	Resilience	I feel full of energy when I am at school.
		I feel bubbly and full of life when I am at school.
		I feel wide awake when I am at school.
		I look forward to coming to school.
	Learning goal orientation	I try hard to do well when I am in class.
		I pay attention when I am in class.
		I listen carefully when I am in class.
	Social harm	I have been verbally harassed.
		I have been physically harassed.
		I have been bullied online or on social media.
		I have been picked on.
Behaviour	Disruptive behaviour	I have used devices in class for non-educational purposes.
		I have been disrespectful to teachers.
		I have misbehaved in class.
		I have ignored the teachers' instructions.
	Risky behaviour	I smoked cigarettes/e-cigarettes or have vaped.
		I drank alcohol.
		I used drugs.
		I skipped lessons or chose not to come to school.
		I damaged school property on purpose.

Students described their outcomes by responding to statements using a five-point Likert scale. The applicability of numerical techniques in analysing this data depends on whether the interval between the responses can be considered equal and, therefore, represents a continuum. The survey (except Institutional Environment) has been comprehensively validated in Australian schools (Aldridge & Ala'I, 2013, p. 59). Importantly, this validation includes confirming that the *outcome* measures can be treated as measuring equally spaced points on a continuum. As such, this data, while ordinal, can be treated as numerical. Treating the data as numerical means that techniques including those used in descriptive statistics are suitable when conducting EDA to better understand student outcomes.

The intent of this research is to use student outcome as the target for predictive analysis. Each student's outcome is defined by their responses to the 28 outcome statements, which are organised into seven separate sub-constructs (see Table 3.4). Predictive analytical techniques exist that are suitable for use with multi-dimensional targets; however, these typically require very large datasets, particularly if the relationships between the input variables and the target are non-linear. More commonly, predictive analytical techniques use a single target variable. Considering the size of the datasets available for each school, this research uses a single target variable.

Two main options are available for implementing single target predictive analysis when faced with multi-dimensional target variables. The first is to produce a model for each target variable and the second is to use dimensional reduction to produce a single target variable. In this case, a hybrid option is also available where the 28 questions are reduced to the seven validated sub-constructs.

Developing a model for each outcome is possible but this would treat the outcomes as independent of each other, which is inconsistent with the complex and dynamic nature of schools. Advanced dimensional reduction techniques such as t-Distributed Stochastic Neighbour Embedding (t-SNE) and Principal Component Analysis (PCA) are capable of reducing the student outcome dataset to a single dimension. However, in enabling computational efficiency, dimensional reduction removes detail from datasets (Wu & Coggeshall, 2012).

In lieu of traditional dimensional reduction, this study uses unsupervised clustering, which is a method that allows the data itself to determine how groups, or clusters, are formed without relying on predefined attributes or categories set by the researcher. This approach ensures that the clusters reflect the data structure without external bias or assumptions. By allowing the data to define the groupings, the analysis can capture subtle or unexpected patterns that may not align with initial assumptions. For example, clusters may emerge based on diverse wellbeing and behaviour indicators, offering a more data-driven and insightful representation of the relationships in the data (Kotu & Deshpande, 2015).

Organising students with similar outcomes into clusters minimises redundancy and refocuses the analysis on fewer, more meaningful groups. These clusters serve as prediction targets in later stages of analysis. Each cluster summarises its members through a representative nominal value (cluster), preserving essential information while reducing complexity. Researchers have previously achieved dimensional reduction using clustering techniques to explore student health outcomes (Wong, et al., 2022) and academic performance (Dumuid, et al., 2017).

Clustering relates to another machine learning method: classification. Classification techniques include logistic regression, decision trees, random forests, and support vector machines. Classification techniques were not considered for this task as they require user-defined categories, whereas clustering analysis was applied for the very purpose of producing the categories upon which these techniques rely. Classification techniques are used in the later *predictive* analysis, as explained in Section 3.2.8.

The study uses X-means clustering, which is a method developed to address a fundamental limitation of the widely used k-means approach. K-means requires researchers to specify the number of clusters (k) before analysis, which can result in too many or too few clusters (Mughnyanti, et al., 2020). X-means eliminates this limitation by defining the number of clusters based on the analysis process (Pelleg & Moore, 2000). The analysis begins with a user-defined minimum number of clusters (k_{min}). The algorithm iteratively splits clusters into two and evaluates whether the resulting model with k_{min+1} clusters improves accuracy. This evaluation relies on statistical criteria such as the Bayesian Information Criterion (BIC), which balances model fit against complexity. The process continues until additional splits fail to enhance the model's accuracy.

Each cluster receives a unique (nominal) label, which the analysis appends to the dataset as a new variable. This cluster label represents the outcome for each student and functions as the target variable in predictive analysis. The predictive models aim to explain these outcomes using the 73 School Climate attributes, transforming raw data into manageable, meaningful clusters for further exploration.

3.2.7 Student experience: Climate misfit and infrastructure experience

The School Climate construct comprises Safety, Interpersonal Relationships, Teaching and Learning, and Institutional Environment. For all but Institutional Environment, the students respond regarding both their current and preferred experience, with the calculated *misfit* field representing the difference. Aldridge et al. (2024) found the relationship between *misfit* and outcomes, specifically resilience, wellbeing and bullying, to be statistically significant. Further, Aldridge et al. found the difference in *misfit* could "provide a more nuanced understanding of students' needs … when compared with relying on the actual data alone" (2024, p. 423). To calculate misfit, the responses to these questions were imported and another attribute was generated by using the difference between them. These misfit responses are identified in the data by the addition of the label "Delta".

This two-step questioning process was not applied when developing the Institutional Environment (infrastructure) questions; the data collected described the students' current experience with infrastructure rather than the *misfit*. There are two reasons for this. First, the Institutional Environment questions represented an additional 23 questions to an already lengthy survey and the NSW Department of Education did not support this. Second, the source questions for the Institutional Environment pillar (detailed in Table 3.1) were developed, tested, and validated as single questions only. As a result, the approach used to measure the student experience with the Institutional Environment is different to that used for Safety, Interpersonal Relationships, and Teaching and Learning. To address this issue, consideration was given to using the response to the *current* question from each comparative pair; however, Dhar et al. (1999) found that current and preferred rate responses cannot be later treated as independent. Hulland et al. (2018) further caution against independent consideration of paired questions as it ignores the inherent systemic bias created by the paired structure. Of the 73 School Climate attributes, 50 were derived calculating the difference between the current and preferred state, these covered Safety, Interpersonal Relationships, and Teaching and Learning. The 23 Institutional Environment attributes reflected student responses to the current state only.

The calculation of misfit resulted in scores between zero and four, whereas the Institutional Environment responses ranged from one to five. For consistency, *normalisation by shifting* aligned the two datasets. This was achieved by increasing all School Climate misfit scores by one, resulting in a consistent scoring scale for all attributes.

3.2.8 Predictive analysis: Overview

This study uses predictive analytics, specifically *classification*, to identify and quantify the relationship between individual student outcomes (wellbeing and behaviour) and their school experience. Predictive analytics has two primary branches: numerical prediction and classification (Larose & Larose, 2015). Numerical prediction estimates continuous variables, and its objective is to minimise the difference between predicted and actual values. In the context of student outcomes, numerical prediction would be an appropriate technique if the objective was to predict student examination results. Classification assigns observations to distinct categories. It supports predictions of categories of outcomes, such as whether students are likely to pass or fail an examination.

Predictive analysis aims to produce a model that can explain the relationship between a student's school experience, as defined by their answers to the 73 School Climate questions, and their outcome group (or cluster). Predicting outcomes based on responses to a School

Climate survey constitutes a sophisticated machine learning task. This task necessitates an algorithm that balances high predictive accuracy with the capacity to model complex interactions among the different School Climate pillars while ensuring interpretability to highlight the predictors of student outcomes. An analysis of several popular classification techniques follows, including the justification for selecting GBTs for this application.

3.2.9 Predictive analysis: Techniques

Decision trees provide a transparent approach to classification by recursively partitioning data based on feature values, creating a hierarchical tree structure where branches represent decision rules and leaves denote predicted outcomes. The graphical representation of this method enables users to trace the reasoning behind the predictions. Additionally, decision trees model both linear and non-linear relationships, demonstrating versatility. However, decision trees tend to overfit, particularly in noisy datasets, which compromises generalisation. Specifically, overfitting occurs when models learn not only the underlying patterns in the training data but also the noise, random fluctuations, and even outliers. The resulting models are too complex, resulting in significant differences between training and test performance (Van Claster, et al., 2019, p. 2). Pruning methods can mitigate overfitting; however, they may introduce subjectivity and degrade performance on complex datasets (Tan, et al., 2019, p. 145).

Logistic regression is a well-established statistical method for binary classification. It models the relationship between predictors and the probability of an outcome through the logistic function, providing interpretable coefficients that quantify the influence of each feature. Logistic regression is computationally efficient and works effectively for datasets where the relationships approximate linearity. However, this technique struggles with non-linear or complex interactions.

Feature engineering, the process of creating or transforming variables to improve the predictive power of machine learning models, offers a valuable tool for enhancing accuracy. However, in exploratory analysis using GBTs, the creation of engineered features inevitably reduces interpretability. Feature engineering combines or abstracts information from the original data, which makes it challenging to trace model predictions back to specific, comprehensible relationships in the raw variables. This transformation obscures the context of the original data because engineered features typically replace straightforward variables with derived representations, such as interaction terms or logarithmic transformations. Although these adjustments improve accuracy, they complicate understanding by distorting the clear connections between inputs and outputs.

The reliance of GBTs on ensembles of small decision trees amplifies this problem. These trees iteratively prioritise engineered features, layering abstraction upon abstraction and further obscuring the role of the original variables. As a result, feature engineering directly obscures relationships in the data, making them harder to interpret and reducing the model's ability to reveal meaningful insights. Exploratory analysis, which focuses on uncovering clear and actionable patterns, suffers significantly in this context. Feature engineering emphasises optimising prediction accuracy, which undermines the model's value as a tool for identifying transparent relationships within the data.

This perspective aligns with findings in the literature. For example, Guarido (2018) demonstrated that while feature engineering improved prediction accuracy in GBTs, it also introduced complexities that could obscure interpretability. Similarly, Delgado-Panadero et al. (2022) discussed the challenges in clarifying the outcomes of GBT models due to their additive nature, which becomes more pronounced with extensive feature engineering. These studies highlight the trade-off between accuracy and interpretability when applying feature engineering in GBT models, particularly in exploratory analyses.

Rule induction generates explicit if–then rules from data, providing easily interpretable models. Nevertheless, this technique proves ill-suited for datasets with overlapping class boundaries or complex feature interdependencies. The fixed nature of rules makes rule induction less effective for tasks intended to capture the complex ways different factors interact and influence the outcome (Larose & Larose, 2015, p. 203).

Naïve Bayes classifiers apply Bayes' theorem while assuming conditional independence among features, meaning that each feature is considered independent of others given the class label. This assumption simplifies real-world dependencies and enables computational efficiency and scalability, particularly when handling large datasets (Manning, et al., 2008). However, this study requires methods suitable for analysing complex relationships, negating the Naïve Bayes approach.

Artificial Neural Networks (ANNs) are powerful tools for modelling complex, non-linear relationships. Their architecture enables them to identify intricate patterns within the data, rendering them suitable for large-scale applications. However, the black-box nature of ANNs limits interpretability, an essential requirement in educational contexts where understanding the drivers of predictions remains crucial. Additionally, ANNs require significant computational resources and careful hyperparameter tuning, diminishing their practicality for this task (Wu & Coggeshall, 2012).

Support vector machines (SVMs) are machine learning models that divide data into categories by finding the optimal boundary (called a hyperplane) that separates the groups. If the data is not easy to separate with a straight line, SVMs use special tools called kernel functions. These functions transform the data into a higher-dimensional space, making it possible to find a boundary that separates the groups more effectively. While effective for high-dimensional datasets, SVMs suffer from limited interpretability. Their computational costs and the implicit nature of their decision boundaries render them less suitable for applications requiring transparency and scalability (Tan, et al., 2019).

GBTs construct decision trees sequentially, with each tree correcting the errors of its predecessors. This iterative approach enables GBTs to effectively model complex interactions among the survey dimensions, such as how low scores in Safety and Interpersonal Relationships interact to predict negative outcomes. GBTs capture linear and non-linear relationships, ensuring accurate predictions even within intricate datasets (Larose & Larose, 2015, p. 275). While GBTs are not inherently easy to interpret, numerous methods exist to enable the identification of important predictive factors and their contribution to outcomes.

Classifying students into positive or negative wellbeing and behaviour outcomes based on School Climate survey responses requires a machine learning model that balances predictive accuracy, complexity, and interpretability. Decision trees and logistic regression provide transparency but struggle with complex relationships. Rule induction and Naïve Bayes classifiers lack flexibility and are ill-suited for handling interdependent features. ANNs and SVMs handle intricate patterns but are less interpretable and scalable. GBTs provide a superior balance, effectively capturing complex interactions, while modern techniques enhance their interpretability, making them the most suitable choice. As such, GBTs are the classification method selected for this research.

3.2.10 Gradient boosted trees: Description and justification

GBTs are a robust machine learning algorithm within the family of ensemble methods. Ensemble learning combines multiple models—often referred to as "weak learners", which perform only slightly better than random guessing—into a more robust predictive model (Larose & Larose, 2015). In regard to GBTs, the weak learners are decision trees. These trees are trained sequentially, with each tree attempting to correct the residual errors of the previous trees. This iterative approach enables GBTs to model complex, non-linear relationships between predictors and outcomes, making them particularly well-suited for datasets with diverse variables.

The GBT algorithm enhances prediction accuracy through an iterative process that corrects errors from previous steps. It starts with an initial prediction, typically the most frequent category in classification tasks. Hastie et al. (2000) explained that this initial guess might need to be adjusted using a logarithmic transformation, which stretches the scale of probabilities to make it easier for the algorithm to work effectively and make better predictions as it learns from the data.

The algorithm computes pseudo-residuals at each iteration, quantifying the discrepancies between current predictions and actual outcomes. These pseudo residuals highlight areas requiring improvement and indicate the steepest gradient for minimising the loss function (Natekin & Knoll, 2013). A newly constructed decision tree addresses these residuals. Unlike more complex models, decision trees partition data based on recognisable patterns. The algorithm employs a scaling factor known as the learning rate to balance the learning pace and mitigate the risk of overfitting (Pezoulas, et al., 2024). The process continues for a predetermined number of iterations or until improvements in predictions stabilise. The final model consolidates the predictions of all decision trees, assigning appropriate weights to each one.

As illustrated in Figure 3.4, the training process starts by building an initial decision tree model that generates predictions based on the input features (Nhat-Duc & Van-Duc, 2023). The first tree (furthest left) provides a baseline approximation of the target values; however, it often produces predictions characterised by substantial residual errors. These residual errors represent the discrepancy between the observed values and the predictions generated by the initial tree.

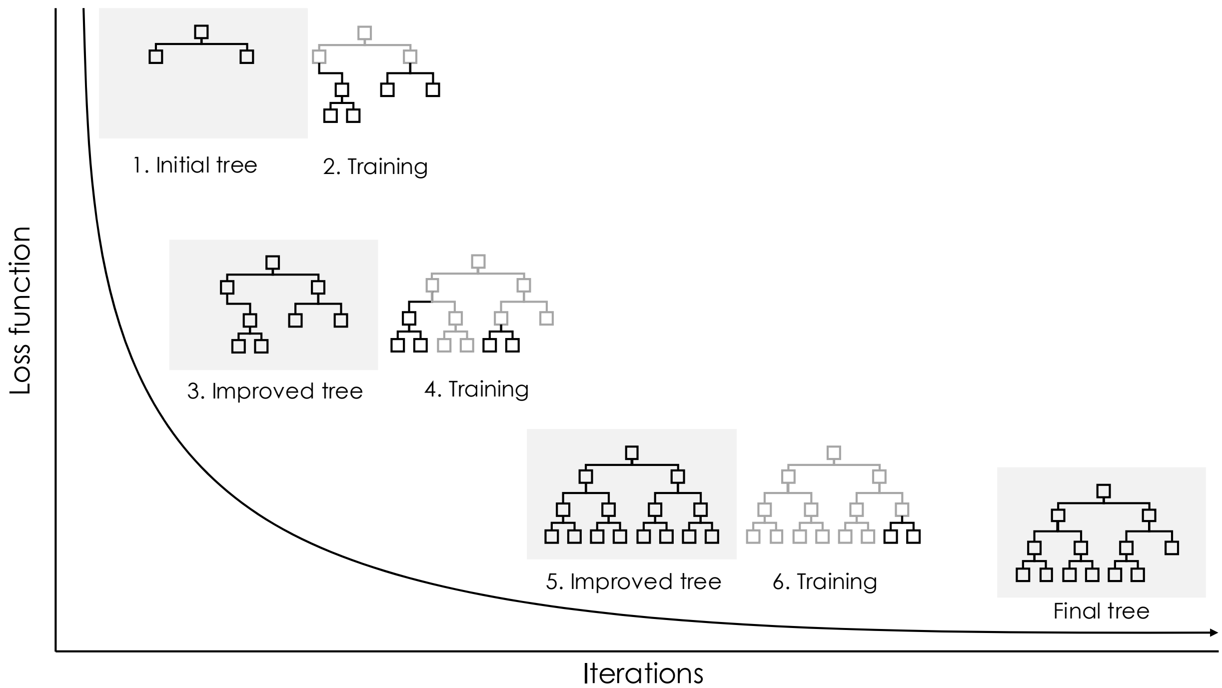

Figure 3.4. The GBT training process

Subsequent phases train a new decision tree to predict these residual errors. The model learns from its previous inaccuracies and incrementally diminishes the prediction error by concentrating on the residuals. This new tree integrates into the existing model, enhancing its predictive performance (Wu & Coggeshall, 2012).

The iterative nature of this process continues, with each additional tree trained to address the residual errors resulting from the predictions of the preceding trees. As the model incorporates more trees over time, it increasingly captures complex patterns inherent in the data. The cumulative effect of this iterative refinement leads to a model that converges towards minimising the loss function, as illustrated by the downward-sloping curve in Figure 3.4 as the number of iterations increases.

The rightmost section of the diagram exemplifies the ensemble of trees generated through this process. Each tree contributes a specific fraction of the final prediction and the aggregated model effectively balances complexity with predictive accuracy (Xu, et al., 2014).

This systematic approach underpins the efficacy of GBT models in addressing both regression and classification tasks. They yield high predictive power while mitigating overfitting through the application of regularisation techniques.

3.2.10.1 Gradient boosted trees: Application in educational research

A growing body of research demonstrates the effectiveness of GBTs in predicting student outcomes. While applied in different circumstances to this study, they have nonetheless been shown to be useful at the (individual) student level in educational environments with their inherent complexity. Martins et al. (2021) applied GBTs to analyse dropout risks at the Polytechnic Institute of Portalegre in Portugal, focusing on those students identified as likely to withdraw from courses early in their academic journey. The study used demographic, academic, and institutional data to demonstrate that GBTs outperformed traditional machine learning models such as Random Forests and SVMs in predictive accuracy and recall. The research also highlighted the flexibility of GBTs in handling imbalanced datasets, a common issue in educational contexts where the number of students in at-risk categories typically constitutes a minority.

Another example is the work by Benkhalfallah and Laouar (2023), which employs GBTs to forecast final examination scores in secondary schools. Their study leverages data on attendance, prior academic performance, and behavioural indicators to identify students likely to underperform. The researchers found GBTs particularly suited to capturing input interactions that influence outcomes.

In addition to predicting student outcomes, GBTs appear in other areas of educational research, such as identifying critical drivers of academic achievement. Farhood et al. (2024) compared GBTs with traditional machine learning models such as Decision Trees and Logistic Regression and found that GBTs consistently outperformed these alternatives in precision and recall. Their study analysed university students' performance and identified attendance, prior academic achievement, and socioeconomic factors as critical predictors. These researchers used the predictive weights generated by the GBTs to provide actionable insights into the factors driving success. These results correspond to those of Tin et al. (2024), who found GBTs to be superior to other models such as logistic regression, Naïve Bayes, Generalised Linear Model, Decision Tree, and VSM in predicting academic performance.

Villar and Andrade (2023) applied GBTs to examine predictors of student engagement in online learning environments. Using behavioural and interaction data from virtual platforms, the researchers found that GBTs were uniquely capable of identifying non-linear relationships between inputs (such as login frequency and time spent on assignments) and outcomes (reported student satisfaction). These examples illustrate the suitability of GBTs for addressing the complex relationships involved in predicting student outcomes and the factors that most influence them.

GBTs provide several advantages that make them well-suited to this study. First, as mentioned previously, GBTs have accurately predicted student outcomes in diverse studies. Second, GBTs effectively model complex, non-linear relationships. Third, GBTs are interpretable despite being an ensemble method. Tools such as feature importance rankings quantify the relationships between predictors and classification outcomes, generating actionable insights for stakeholders.

3.2.10.2 Gradient boosted trees: Training

GBTs are a learning method and, as such, must be trained. Training involves splitting the available data into training and testing sets. GBTs are also an ensemble method that iteratively improves the results of weak trees to achieve incremental improvement, directed by user (researcher) defined parameters. This section explains the training process and important decisions, including the proportion of data assigned to training and testing and the parameters guiding the learning process (Yu & Ho, 2022).

The task of the GBTs in this research is to predict the outcome cluster (positive or negative) to which a student belongs, as assigned by the cluster analysis. The input data for the prediction is the School Climate survey, recording a student's school experience. For training, a GBT uses an input (School Climate) and an output (Outcome Cluster), with the intent to learn the predictive relationships between the student's school experience (School Climate) and their wellbeing and behaviour (Outcome Cluster). The GBT is provided with unseen input data (School Climate responses) for testing and to predict the output (Outcome Cluster).

Splitting the data into training and testing sets is fundamental to predictive analysis using GBTs and other learning methods. The intent is to ensure sufficient training data so that the resulting model (GBT) can effectively predict unseen data. There is no universally accepted optimum ratio, with the choice being influenced by factors such as the heterogeneity of the data and the size of the overall dataset.

To understand the effects of data size (response rate) on survey prediction, a single ratio of 70:30 (training to testing split) was chosen for all schools and the whole-of-cohort. According to Touzani et al. (2018), this split provides a practical balance by allocating enough data to training for effective learning while maintaining a testing subset large enough to assess generalisation accurately. This ratio has proven effective in diverse applications, particularly for larger datasets with a substantial 70% training subset. For smaller datasets, such as those at the organisational level, a 70:30 split provides a larger testing set than an 80:20 split, although it reduces the size of the training set. The testing set remains essential for evaluating predictive performance, as smaller datasets risk unreliable evaluation metrics due to statistical variability (Yang, et al., 2020). Additionally, smaller datasets are more susceptible to overfitting when the training set is too large, which reduces model accuracy on unseen data.

While this 70:30 split may not yield the largest possible training set for smaller datasets, it represents a deliberate compromise that balances the competing demands of training and testing. Reserving 30% of the data for testing ensures that the evaluation subset remains sufficiently large to provide reliable performance metrics while allocating the majority of the data for training.

With 70% of the data allocated to training, GBTs can typically identify meaningful patterns, even in smaller datasets, provided that appropriate parameters are selected. An important parameter is the learning rate, which determines the contribution of each tree to the model. The learning rate is a critical hyperparameter that requires careful tuning to account for the reduced size of smaller training subsets. A lower learning rate, typically between 0.01 and 0.1, allows the GBT to make incremental adjustments, reducing the risk of overfitting (Arocupita, et al., 2024). However, lower learning rates necessitate more boosting iterations to achieve convergence, which can increase computational costs. As this research aimed to understand the predictive relationships, the computational cost—while important for deployed production models—was not deemed important in this research. As such, the chosen learning rate was 0.01.

The maximum depth of trees is another key parameter that influences the model's capacity to capture complex relationships in the data. For the whole-of-cohort dataset, tree depths of three to five typically balance model complexity and the risk of overfitting (Touzani, et al., 2018). This depth range remains applicable in smaller organisation-level datasets, but monitoring is essential to ensure that the model does not overfit the limited training data. For this research, the maximum depth was set at five.

The number of trees, or boosting iterations, further supports the refinement of predictions. While larger datasets can benefit from 300 to 500 trees, smaller datasets may require fewer iterations, typically 100 to 300, to prevent overfitting and maintain computational efficiency (Yang, et al., 2020). For this task, the number of trees was set to 300, striking an appropriate balance for both small and large datasets and ensuring robust performance without excessive computational cost.

The main parameters used in optimising GBTs are the split between training and testing datasets, the learning rate, tree depth, and the number of trees. Table 3.4 presents all the parameters and the values chosen.

Table 3.4. GBT parameters and values

Parameter	Value
Training to testing split	70:30
Number of trees	300
Maximum tree depth	5
Learning rate	0.01

In conclusion, the consistent use of a 70:30 split represents a reasoned compromise that balances training and testing needs across datasets of varying sizes. While smaller datasets might benefit from alternative splits that allocate more data to training, the 70:30 ratio ensures a sufficiently large testing subset to produce reliable evaluation metrics. For this research, a maximum depth of five was selected, aligning with the recommendations from Touzani et al. (2018) for both large and small datasets. Similarly, 300 trees were chosen, providing sufficient predictive power while mitigating the risk of overfitting, as recommended by Yang et al. (2020). These parameter choices support practical model training and evaluation for predicting student outcomes.

3.2.10.3 Gradient boosted trees: Evaluating performance

A confusion matrix is an essential tool for assessing the performance of a classification model, offering a comprehensive breakdown of its predictions. It classifies these predictions into four categories: true positives, false positives, true negatives, and false negatives. True positives occur when the model accurately identifies a present condition or class. In contrast, false positives arise when the model erroneously predicts the presence of a condition that does not exist, indicating instances where it mistakenly identifies something as positive. True negatives represent cases in which the model correctly predicts the absence of a condition, affirming that it is indeed not present. Finally, false negatives occur when the model fails to detect a present condition, omitting instances it should recognise (Wu & Coggeshall, 2012).

This idea is illustrated in Table 3.5, with two clusters, A and B, representing the positive and negative clusters, respectively.

Table 3.5. Confusion matrix for clusters A and B

Confusion matrix		**Actual class (observation)**	
		Cluster A	Cluster B
Predicted class (expectation)	Cluster A	Prediction = A Actual = A TRUE Positive (TP)	Prediction = A Actual = B FALSE Positive (FP)
	Cluster B	Prediction = B Actual = A FALSE Negative (FN)	Prediction = B Actual = B TRUE Negative (TN)

The data in the confusion matrix allows for four different assessments of model "goodness": sensitivity, specificity, precision, and accuracy. These measures are defined and the calculations specified in Table 3.6.

Table 3.6. The four measures of model performance (Kotu & Deshpande, 2015, p. 260)

Term	**Definition**	**Calculation**
Sensitivity	The ability of the model to identify instances in Cluster A.	TP/(TP+FN)
Specificity	The ability of the model to avoid misclassifying instances from Cluster A as Cluster B.	TN/(TN+FP)
Precision	The ability of the model to avoid misclassifying instances from Cluster B as Cluster A.	TP/(TP+FP)
Accuracy	Directly reflects the overall correctness of the model, aligning with the goal to classify without bias.	(TP+TN)/(TP+TN+FP+FN)

3.2.10.4 Receiver Operating Characteristic Curve and Area Under the Curve

The Receiver Operating Characteristic (ROC) curve represents the trade-off between *true* and *false positive rates* across different classification thresholds (Fan, et al., 2006). It is particularly useful for evaluating the model's ability to distinguish between classes independently of any specific threshold (Wu & Coggeshall, 2012, p. 222).

The Area Under the Curve (AUC) quantifies the ROC curve's overall performance, with values ranging from 0.5 (no discrimination) to 1.0 (perfect discrimination). Benchmarks for interpreting AUC scores categorise performance from unsatisfactory to excellent. Thresholds define scores below 0.6 as unsatisfactory, indicating poor model performance. Scores from 0.6 to 0.7 reflect satisfactory performance, meeting minimal standards. Scores from 0.7 and 0.8 indicate good performance, demonstrating moderate predictive capability. Scores from 0.8 to 0.9 represent very good performance, reflecting substantial predictive accuracy. Finally, scores above 0.9 indicate excellent predictive accuracy (Trifonova, et al., 2013). Figure 3.5 illustrates the idea of AUC.

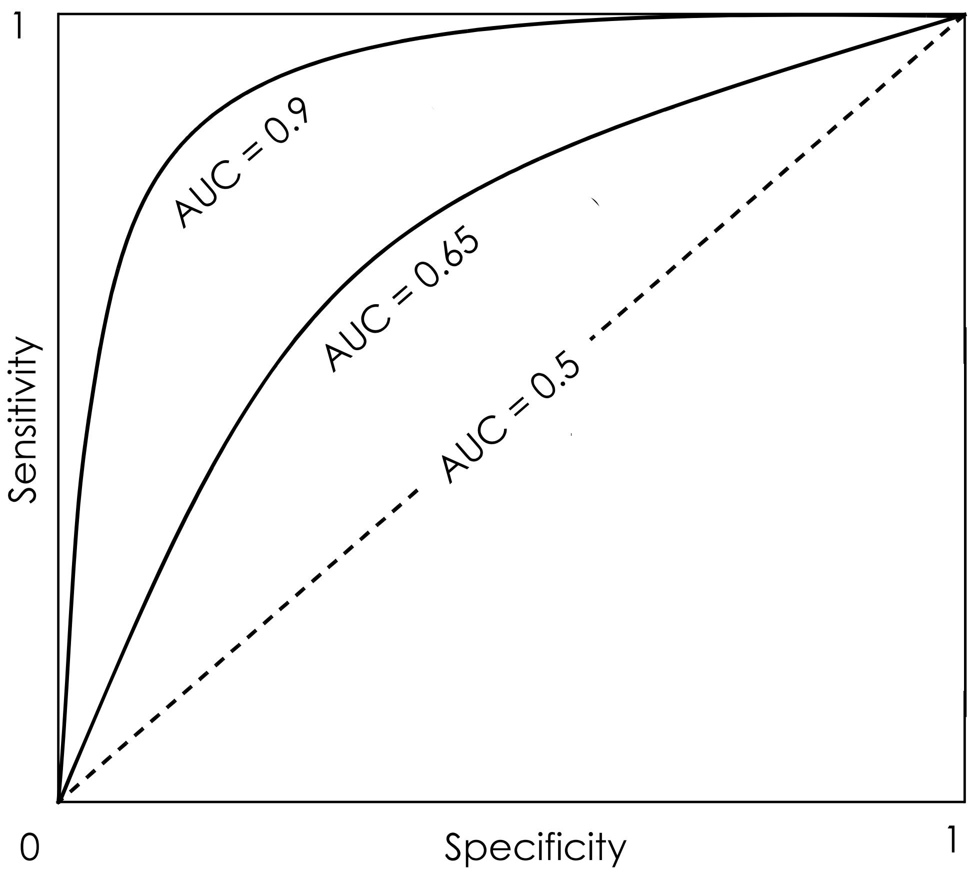

Figure 3.5. ROC curves with good (AUC = 0.9) and satisfactory (AUC = 0.65) parameters of sensitivity and specificity (adapted from Trifonova et al., 2013, p. 182)

These benchmarks provide a standardised framework for evaluating model performance and are particularly useful for comparing the effectiveness of different classifiers in distinguishing between positive and negative instances. The AUC analysis complements the confusion matrix by providing an aggregated view of model performance across thresholds, making it a critical tool for evaluating the robustness of the GBT model (Fawcett, 2006).

3.2.10.5 Gradient boosted trees: Extracting insights

GBTs are widely recognised for their ability to model complex, non-linear relationships between predictors and outcomes. These algorithms iteratively combine multiple weak learners, typically shallow decision trees, to improve predictive accuracy. By correcting residual errors in successive iterations, GBTs excel at capturing intricate patterns in the data, making them highly effective for tasks such as predicting student outcomes.

In this study, where multiple demographic, institutional, and behavioural factors interact dynamically, GBTs are particularly appropriate. However, their ensemble-based structure, which relies on sequentially trained decision trees, makes them inherently difficult to interpret. Compared to simpler models, such as logistic regression or single decision trees, the aggregation of numerous trees obscures the relationship between predictors and outcomes, complicating efforts to understand the drivers behind predictions.

Interpretability techniques such as predictive weights (Kotu & Deshpande, 2015) and SHapley Additive exPlanations (SHAP) (Ibrahim, et al., 2022)address the opacity of GBTs. Both techniques aim to explain how features contribute to model predictions; however, they differ significantly in their methodology and suitability for specific research objectives. In this study, predictive weights serve as the primary interpretability method due to their alignment in analysing global patterns across ten schools, which are complex and dynamic systems.

Predictive weights provide a global measure of feature importance by quantifying each predictor's contribution to the overall performance of the GBT model. These weights are calculated by aggregating information from all decision trees in the ensemble. Specifically, they assess how frequently a feature is used for splitting nodes, the depth of these splits, and the reduction in the loss function achieved when the feature is employed (Xu, et al., 2014)

Features that frequently contribute to reducing the loss function are more likely to be selected near the root nodes of decision trees in gradient boosting models, thereby receiving higher importance scores.. By summarising feature importance across the entire dataset, predictive weights provide a comprehensive overview of which predictors significantly influence the model's outputs. Unlike instance-specific methods, predictive weights do not explain individual predictions; instead, they focus on the global contribution of each feature.

In the context of this study, predictive weights are particularly advantageous for comparing the contributions of predictors across ten schools, each representing a unique educational environment. This approach views schools as complex systems in which diverse factors such as institutional infrastructure, student demographics, and behavioural patterns interact differently.

Predictive weights simplify this complexity by summarising the overall importance of predictors, allowing for the identification of system-wide patterns and cross-school differences. This flexibility makes predictive weights ideal for understanding broad patterns while accounting for the variability inherent in dynamic systems such as schools (Martins, et al., 2021).

SHAP, by contrast, is an instance-specific interpretability method that decomposes individual predictions into contributions from each feature. Unlike predictive weights, which summarise feature importance at a global level, SHAP focuses on explaining the marginal impact of each predictor on a specific prediction (Ibrahim, et al., 2022).

While SHAP is effective for understanding individual predictions, its suitability for large-scale analyses is limited. The computational cost of SHAP is high, as it evaluates feature contributions across all possible subsets of predictors. In this study, where the objective is to analyse system-wide trends and cross-school comparisons, predictive weights provide a more practical and aligned approach.

In summary, while predictive weights and SHAP offer valuable tools for interpreting GBT models, predictive weights were selected for this study due to their computational efficiency, alignment with the research objectives, and ability to provide a global perspective on feature importance. Their suitability for analysing complex and dynamic systems ensures that the findings are robust and actionable, supporting the study's aim of identifying the institutional and behavioural predictors that drive student outcomes across the ten schools.

3.2.10.6 Gradient boosted trees: Assessing complexity

As explained in section 3.2.10, GBTs can analyse linear, non-linear, and highly complex relationships between dependent and independent variables, whereas regression techniques are more suited to linear and less complex relationships (Wu & Coggeshall, 2012). As such, when using the same dataset, differences in results achieved by regression analysis and GBTs can reflect their relative abilities in dealing with non-linearity (Hamandan & Ahmad Ganai, 2023, p. 4). The relationships are more likely to be linear when both techniques identify the same (or similar) attributes as being the most predictive. Significant differences between the rankings produced by GBTs and regression analysis are likely attributable to non-linear or complex relationships. This difference in ranking occurs because non-linear and complex relationships are not appropriately identified by regression analysis (Xu, et al., 2014, p. 522), attracting a low (regression-based) rank. The GBTs assign a high predictive weight (and rank) if these non-linear relationships are highly predictive. As such, those attributes that are ranked high by GBTs and low (relative to the GBTs) by regression analysis are of particular interest. A significant motivator in selecting GBTs is the view that schools are complex and dynamic systems. That is, the assumption of complexity justified the additional computational complexity and the need for predictive weights. Assessing the nature of the relationships between School Climate and outcomes is an important step in understanding the drivers of student outcomes.

This study employs GBTs to predict student outcomes and identify the factors that are most important in this prediction. GBTs are selected because they capture complex, non-linear relationships between inputs and outcomes. A suite of performance measures is employed to assess performance across various criteria. Predictive measures are selected to interpret the outcomes and understand the significance of features and the interconnections among predictors.

3.3 Execution

The previous section detailed the EDA, the selection of X-means clustering analysis to define outcome groups and assign respondents to these, and the selection of GBTs as the preferred predictive analysis typology. This section addresses the selection of RapidMiner as the analytical platform and the execution of the analytical process.

RapidMiner, a product from the University of Dortmund (Rapidminer, 2023), was selected as the package for designing and executing the analytical process. Users develop models in RapidMiner by constructing processes using a Graphical User Interface (GUI). RapidMiner employs object-oriented programming, offering both with standard and customisable objects (routines).

Python is a high-level programming language widely used in predictive analytics due to its extensive machine learning libraries, scalability, and strong support for data manipulation and model development. RapidMiner's performance parameters align closely with Python's, indicating comparable efficiency and scalability in data analysis tasks. However, RapidMiner has two desirable features that separate it from Python for this research project. Designed as an enterprise-level solution, RapidMiner has robust traceability and audit features. The "no black box" philosophy records every step taken across the analytical process (Kovács & Ghous, 2020). RapidMiner provides a visual flowchart of each process (Kotu & Deshpande, 2015). As such, any reviewer can quickly interrogate each process step by step. In comparing Rapidminer to other tools (such as Weka, R tool, and Knime), Dwivedi et al. (2016) note that Rapidminer has the most comprehensive suite of algorithms and is the only tool (of those compared) that has both predictive and statistical capabilities.

Students filled out the surveys online using tablet devices supplied by the NSW Department of Education. The department forwarded responses as school-specific CSV files. On receipt, each file was quality-checked. At the import stage, all variables, regardless of type, error rate, or the prevalence of missing values, were imported into the master dataset. Assigning a type to each attribute was essential to the import process. That is, identifying each attribute, or the data used to describe this attribute, using analytical typology. This attribute type dictates the suitable analytical methods.

Figure 3.6 illustrates the process of implementing exploratory data analysis, X-means clustering, and GBTs in RapidMiner. A detailed description of the model architecture, including an explanation of each specific step in executing the RapidMiner is included in Appendix 4.

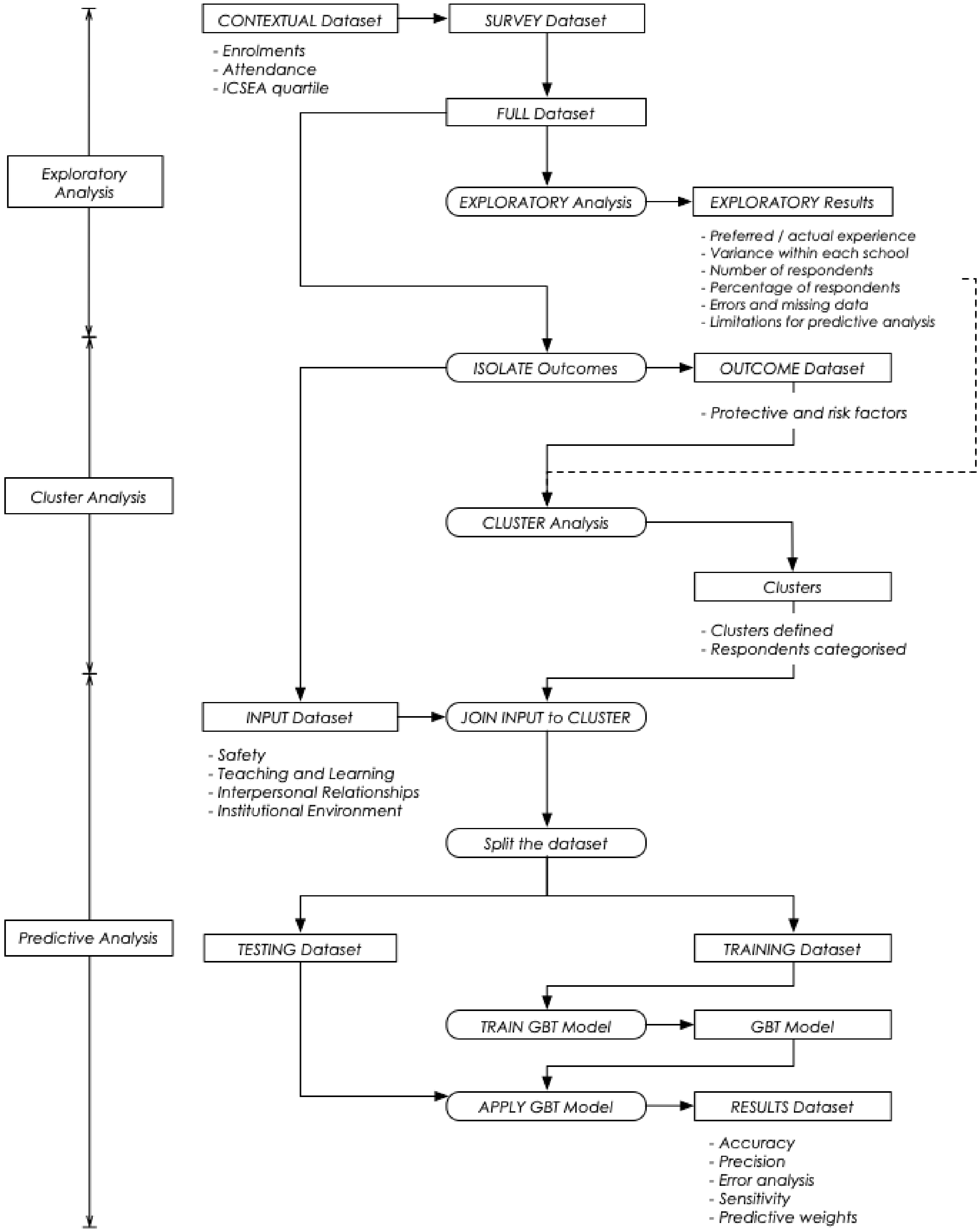

Figure 3.6. Overview of exploratory analysis, clustering, and predictive analysis

This chapter explains the methodology employed to investigate the relationship between school infrastructure and student outcomes, focusing on the development, design, and execution of the research. The pragmatic paradigm underpins the study, which was chosen for its compatibility with mixed methods and suitability for addressing complex systems such as schools. This philosophy bridges objectivist and subjectivist approaches, offering the flexibility to explore the research question effectively. An inductive approach guides the study, enabling theory to emerge from the data and aligning with the exploratory nature of the research.

The data collection used a modified School Climate survey developed by Aldridge and collaborators and extended to include the Institutional Environment construct, which encompasses factors such as maintenance, accessibility, and outdoor spaces. This adaptation, informed by existing literature and stakeholder consultation, ensured that the survey captured the role of infrastructure within the broader School Climate framework. The NSW Department of Education facilitated the survey administration across the ten secondary schools, collecting responses through both electronic and paper formats.

The analysis employed RapidMiner, which was chosen for its advanced algorithms, traceability, and graphical workflow interface. Exploratory data analysis established patterns and assessed data quality, whereas X-means clustering grouped students into outcome categories based on wellbeing and behaviour constructs. GBTs, chosen for their capacity to model complex and non-linear relationships, provided predictive insights into the link between School Climate attributes and student outcomes. Predictive weights quantified the influence of individual predictors, offering actionable insights.

The methodology establishes a robust framework for analysing the dynamic interactions between School Climate and student outcomes. The results chapter will present the findings of these analyses and provide evidence to inform educational decision-making.

Chapter 4. Results

This section details the analytical results. The first section covers the *exploratory analysis*, providing an overview of the data collected, data quality, and descriptive statistics. The second section of this chapter details the results of the X-means clustering analysis, including the generation of the outcome categories used in the predictive analysis. The third section covers the *performance* of the GBTs in predicting educational outcomes. This is important because the value of any insights provided by the GBTs must be weighted according to the underlying predictive performance. Finally, the fourth section details the relative importance of the individual School Climate attributes in predicting educational outcomes. Considering the research question, this fourth section focuses on the importance of Institutional Environment attributes in predicting educational outcomes.

4.1 Exploratory data analysis

This section explores the survey data to uncover patterns, assess data quality, and describe general trends across key domains of the student experience. The analysis covers School Climate and educational outcomes. Descriptive statistics and visual summaries examine how students perceive their current school experience; how it compares to their preferred environment; and how these perceptions align with self-reported behaviours and wellbeing indicators.

The analysis also serves a dual purpose: ensuring the integrity of the dataset and providing foundational insights that inform later inferential and predictive modelling. As such, this section offers a snapshot of the student responses and a framework for understanding how data-driven decisions can be grounded in the students' experiences of the school.

4.1.1 Data quality

The School Climate survey was administered with the intent of eliciting an understanding of school experience in four specific areas: Safety, Interpersonal Relationships, Teaching and Learning, and Institutional Environment. This section of the report examines the responses provided, specifically the shape of the data as described using descriptive statistics.

The dataset, delivered as CSV files, was directly imported into RapidMiner. The data preparation process used RapidMiner packages to check for missing data and type mismatches. Additionally, the data underwent validation to confirm that all responses fell within the ranges specified by the survey questionnaire.

The data covering School Climate and educational outcomes did not contain missing values or outliers. However, contextual questions relating to gender and ethnicity contained some missing and erroneous (non-sensical) values. These responses were not critical for this study and, as such, were not imported. The NSW Department of Education included other questions that also allowed students to provide free-text responses. These questions were irrelevant to this research; therefore, the responses were not imported.

4.1.2 School Climate: Misfit (Safety, Interpersonal Relationships, and Teaching and Learning)

A difference between the *current* and *preferred* experience indicates a climate misfit. This climate misfit, derived from two five-point Likert scales (one measuring *current experience* and the other measuring *preferred experience*), ranges from zero (indicating no difference between current and preferred experience) to four (representing the maximum possible difference). Figure 4.1 details each school's mean School Climate misfit excluding Institutional Environment, that is, the difference between the current and preferred experience of Safety, Interpersonal Relationships, and Teaching and Learning.

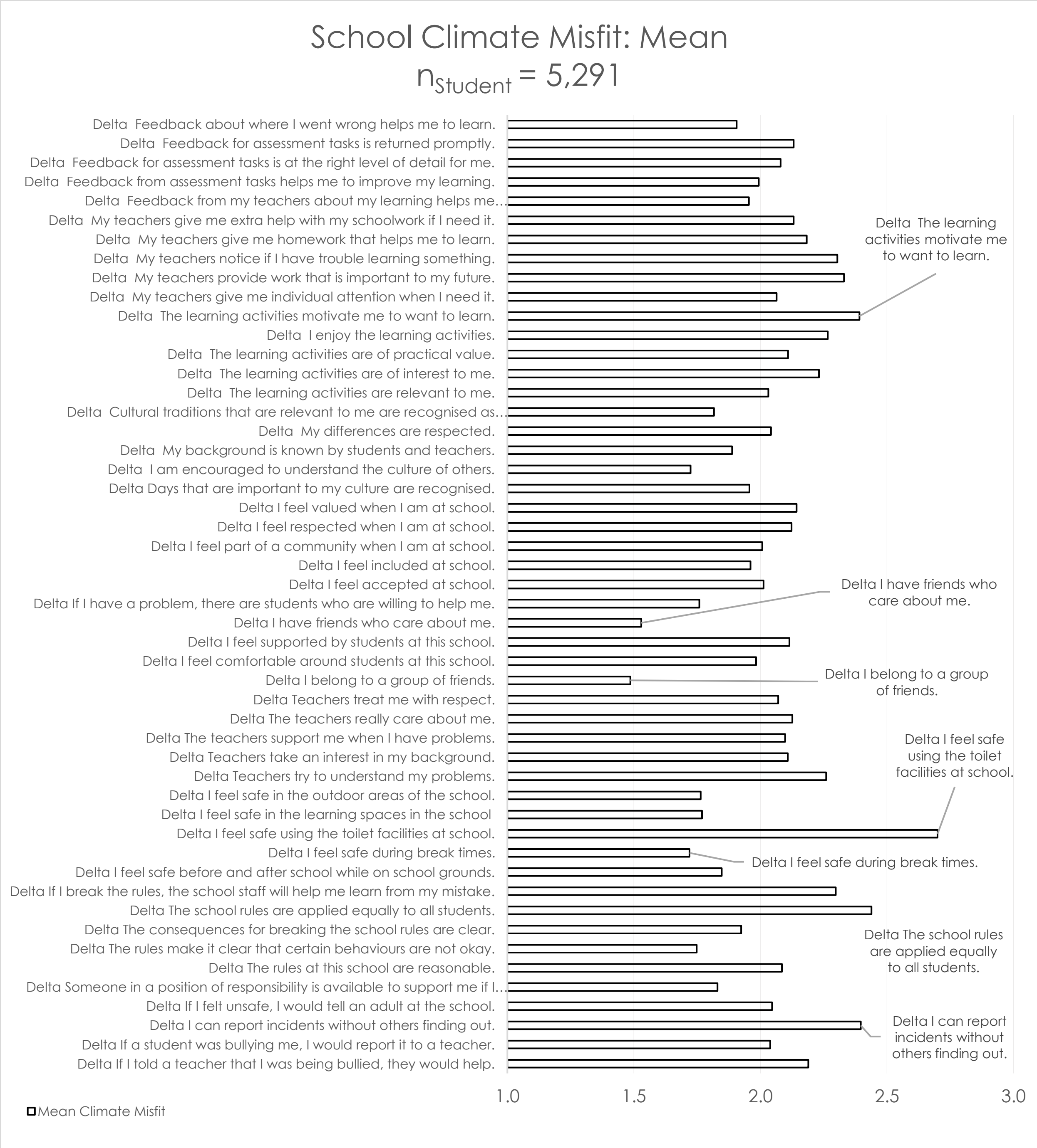

Figure 4.1. Survey results for School Climate misfit

Areas of the highest misfit, indicating the largest gap between preferred and current experience, include feeling safe in the toilets, how much students are motivated to learn by classroom and other learning activities, and the consistency with which rules are applied at their school. Also shown in Figure 4.1 are those areas where the misfit is lowest, indicating elements of school experience where the student's preferred experience is closest to the current situation. These areas include belonging to a group of friends and having friends who care. A chart detailing the descriptive statistics for School Climate Misfit is included in Appendix 5. The misfit scores span the entire range for each attribute, from zero (no misfit) to four (maximum misfit). For each attribute, there are some students for whom the current and preferred experience are the same and others for whom the difference between current and preferred experience is as large as can be represented by the scoring regime.

4.1.3 School Climate: Institutional Environment (infrastructure)

Figure 4.2 presents the mean response data for the Institutional Environment section of the School Climate survey, reflecting the students' experiences with school infrastructure. The Institutional Environment responses are scored from 1 to 5, using a five-point Likert scale ranging from strong disagreement (1) to strong agreement (5). Students responded with their current view of the Institutional Environment only, so no calculation of difference (misfit) was required.

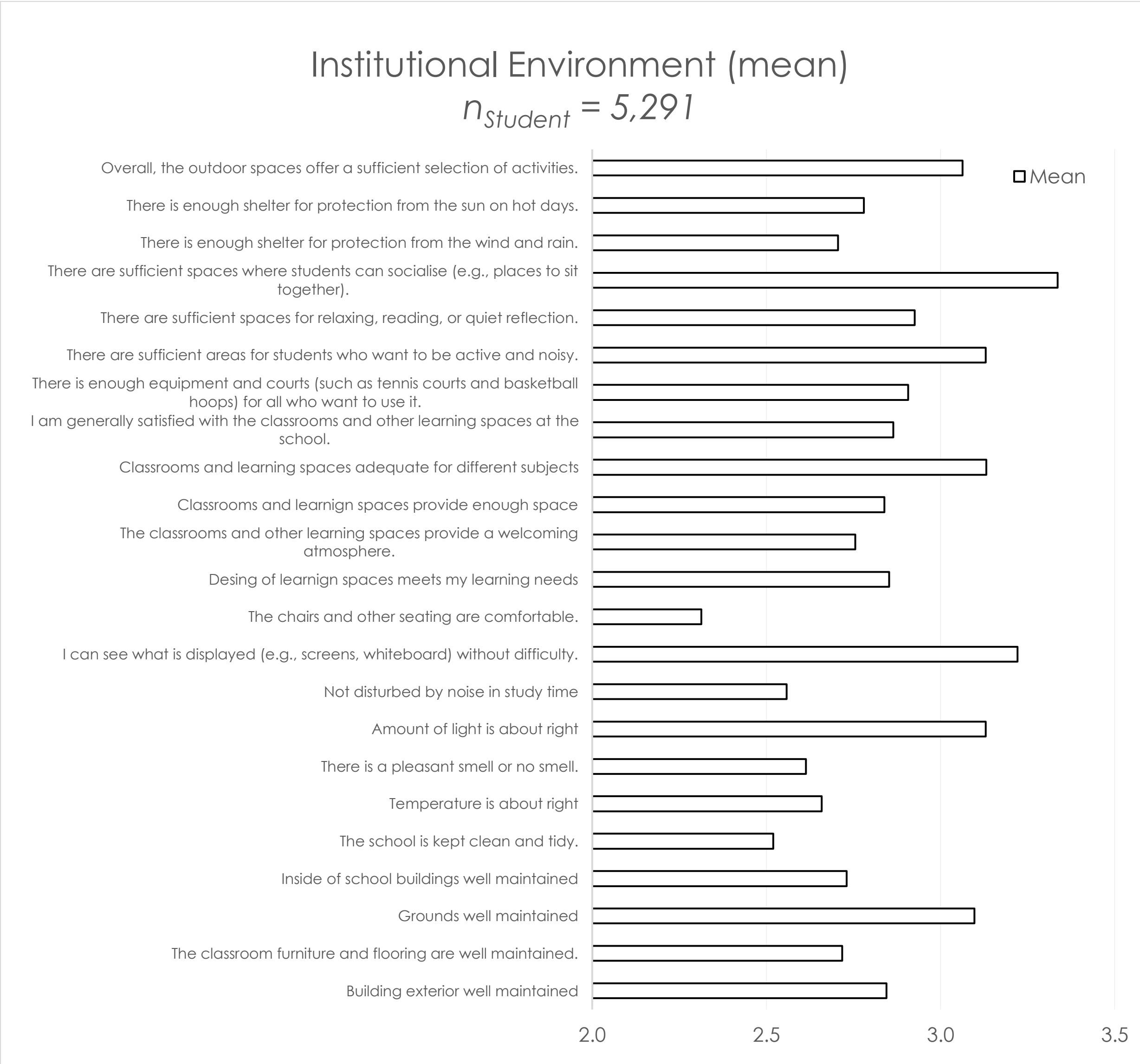

Figure 4.2. School Climate survey results for Institutional Environment

On average, students are least satisfied with their experience with infrastructure in the areas of cleanliness and tidiness, smell, noise when trying to study, and the level of comfort afforded by the chairs and furniture. However, students are most satisfied with their experience of infrastructure in classrooms, visibility of information in lessons, and spaces for socialising and active play.

Descriptive statistics for Institutional Environment responses are included in Appendix 6. As with the responses regarding School Climate misfit, student feedback on Institutional Environment attributes spans the full range of the five-point Likert scale. In this scale, a score of 3 represents a neutral or midpoint response. For most attributes, the distribution of responses is skewed towards the negative, indicating that students tend to disagree, to varying degrees, with the statements provided. Specifically, the attributes that received the highest levels of disagreement include the maintenance of furniture and fittings, the cleanliness and tidiness of the school, temperature, odour, intrusive noise, comfort of furniture, and protection from the elements.

4.1.4 Educational outcomes

The survey required students to respond regarding their school experience (School Climate) and educational outcomes (social and psychological wellbeing and behaviour). In the outcomes section of the survey, four areas contribute to the wellbeing assessment: self-efficacy, vigour, resilience, and learning goal orientation. Behaviour is assessed in three areas: social harm, disruptive behaviour, and risky behaviour. The mean outcome responses for all students are presented in Figure 4.3.

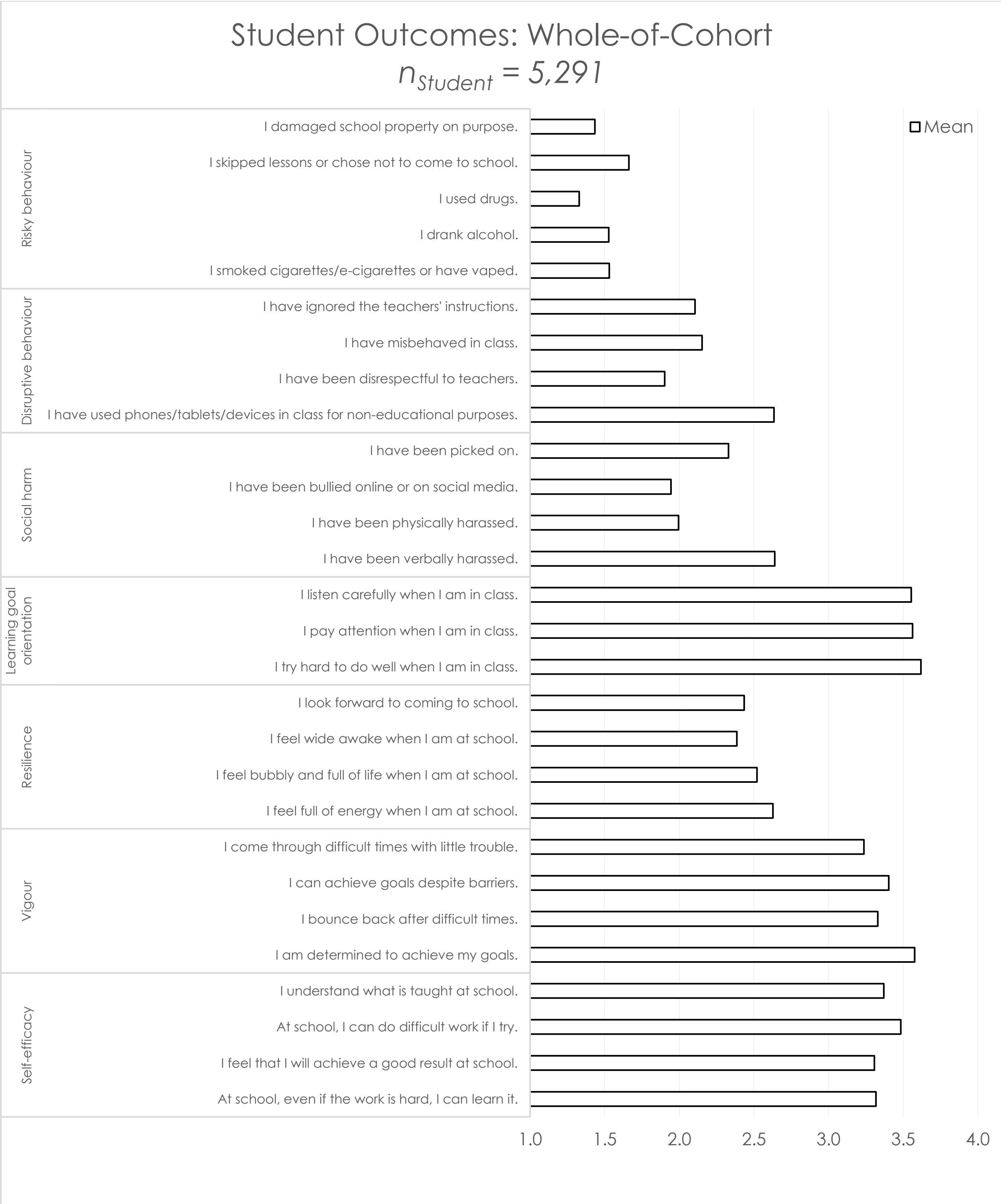

Figure 4.3. School Climate survey results for student outcomes

The results reveal substantial variation in the students' self-reported behaviours and experiences across the various domains. The responses captured using a five-point Likert scale (where 1 = "Always" and 5 = "Never") indicate that students frequently engage in positive academic behaviours and report lower frequencies of risky or disruptive actions.

Students consistently report high levels of academic engagement and motivation. They frequently agree with statements such as "At school, I can do difficult work if I try" and "I am determined to achieve my goals", which receive mean scores approaching 4. Similarly, behaviours related to effort and attentiveness in the classroom, such as "I try hard to do well when I am in class" and "I pay attention when I am in class", also yield high mean scores. These findings suggest that most students perceive themselves as capable and motivated learners, who actively engage in their academic responsibilities.

In contrast, students report experiencing social challenges less frequently. They score statements addressing bullying and harassment, such as "I have been physically harassed" and "I have been bullied online or on social media", closer to 1.5–2.0, which indicates that while such experiences occur relatively infrequently within the cohort, they are not negligible. The data also indicates that some students admit to behaviours that challenge school norms. For example, students score statements such as "I have been disrespectful to teachers" and "I ignored the teacher's instructions" in the mid-range, reflecting a less consistent, yet still present, occurrence of negative interactions with authority figures.

Behaviours categorised as risky or overtly disruptive, such as "I drank alcohol", "I smoked cigarettes/e-cigarettes or have vaped", and "I used drugs", receive the lowest mean scores, around 1.0–1.5. Similarly, "I damaged school property on purpose" also scores in this range, suggesting that such behaviours appear infrequently within the cohort.

While these findings provide an overall view of student behaviours and experiences, the mean scores likely mask significant variation between subgroups of students. The following section details the findings of the X-means cluster analysis. This analysis was undertaken for two reasons. First, to provide a deeper understanding of how different groups of students share distinct behavioural profiles and experiences, allowing for a more nuanced interpretation of the cohort's outcomes. Second, to provide a target for the GBTs in the predictive analysis stage of the research.

4.2 Cluster analysis

RapidMiner was used to execute X-means cluster analysis to group students based on their reported wellbeing and behaviour. The clusters are defined using a central vector, or centroid, for each outcome attribute. Cluster analysis was undertaken at the school-by-school level and the whole-of-cohort level. The following subsections describe the clusters identified at the whole-of-cohort level; the proportional allocation of clusters at the school-by-school level; and the proportion of students assigned to each cluster.

The X-means clustering identified the same two clusters at both the whole-of-cohort level and at each school. Unlike the k-means clustering approach, the X-means algorithm determines the most appropriate number of clusters based solely on the data provided. Figure 4.4 shows the centroid chart for all students, that is, clustering analysis that treats the students as a single population rather than as separate school-specific populations. The vertical axis represents the responses to the five-point Likert scale, and the horizontal axis represents wellbeing attributes (self-efficacy, vigour, resilience, and learning goal orientation) and behavioural attributes (social harm, disruptive behaviour, and risky behaviour).

The identification of two *outcome clusters*, used to categorise individual students, remains consistent across all schools and the whole-of-cohort. As shown in Figure 4.4, the two clusters represent students who report relatively *positive* and relatively *negative* outcomes. Specifically, the students in the *positive* cluster report higher levels of self-efficacy, vigour, resilience, and learning goal orientation. The students in the positive cluster report relatively lower exposure to social harm and are less likely to participate in disruptive or risky behaviour. By contrast, the students in the *negative* cluster report relatively more exposure to social harm and are more likely to participate in risky and disruptive behaviour. The students in the negative cluster also report lower self-efficacy, vigour, resilience, and learning goal orientation. The X-mean cluster analysis results for each school are included in Appendix 5.

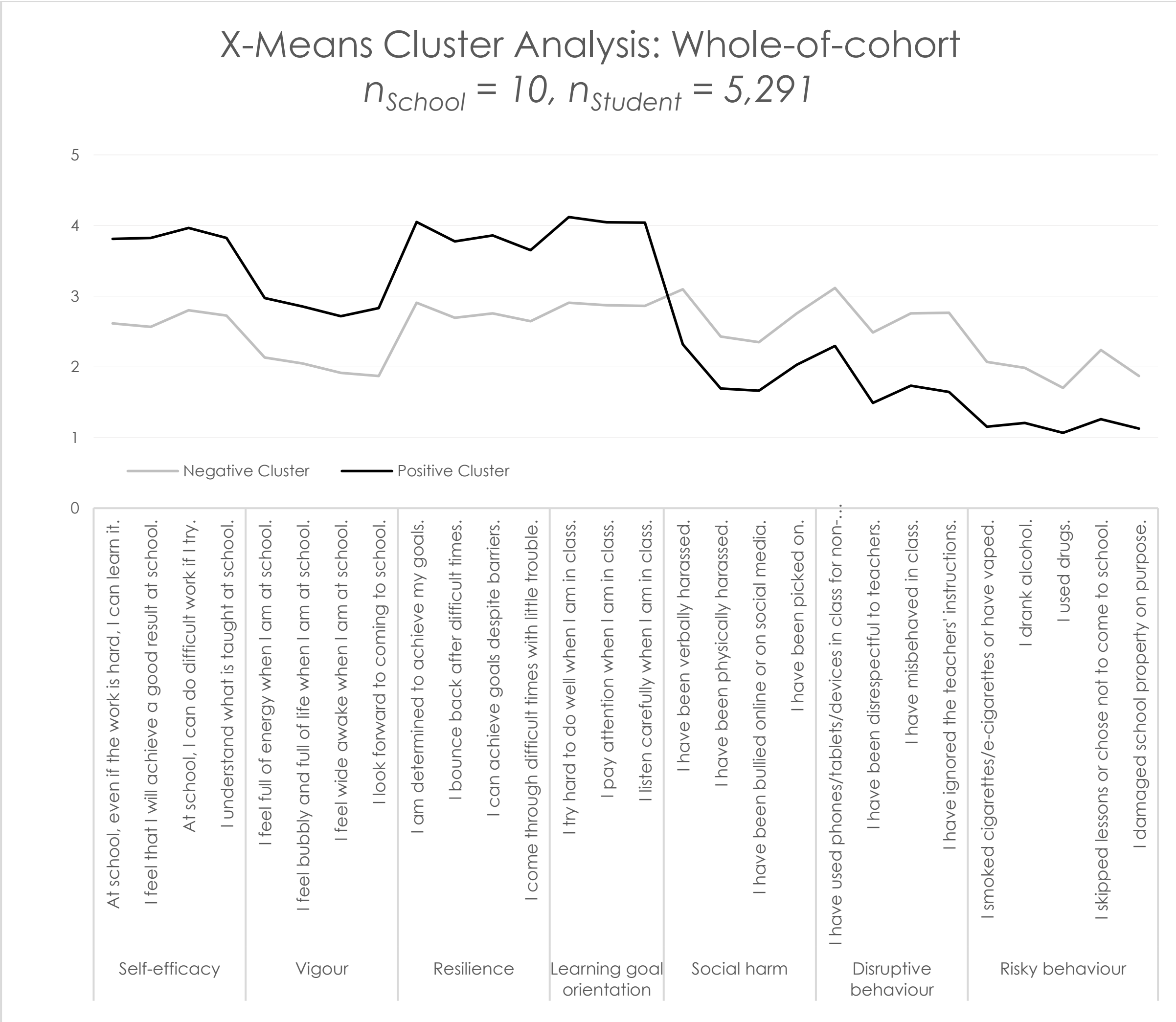

Figure 4.4. X-means clustering results for student outcomes

While the identification of two clusters, positive and negative, is consistent across all the schools, there are, however, differences in the clusters. The cluster analysis for each school is included in Appendix 7.

One measure of difference is how much the two clusters differ between schools as measured by the *range* between the positive and negative centroids for each element. Identifying two clusters, labelled "positive" and "negative", proves consistent across all schools; however, the specific characteristics of these clusters vary considerably depending on the school. Researchers can quantify these differences by examining the distances between the centroids of the two clusters across various attributes. The centroid distances, presented in Table 4.1, illuminate the divergence in student-reported outcomes for each school and enable a comparative analysis across institutions. Each value in the table represents a percentage of the mean distance for the corresponding attribute across all schools, which assists in identifying school-specific patterns. Shaded cells indicate a centroid distance less than 50% of the average, or greater than 200% of the average.

Table 4.1. Centroid distances for each attribute by school

		School									
	Attribute	A	B	C	D	E	F	G	H	I	J
Self-efficacy	At school, even if the work is hard, I can learn it.	82%	102%	102%	108%	66%	115%	119%	96%	105%	105%
	I feel that I will achieve a good result at school.	84%	97%	108%	97%	68%	119%	112%	105%	100%	110%
	At school, I can do difficult work if I try.	99%	94%	110%	105%	49%	108%	119%	112%	100%	105%
	I understand what is taught at school.	91%	96%	110%	102%	72%	110%	117%	106%	97%	101%
Vigour	I am determined to achieve my goals.	71%	78%	106%	140%	59%	99%	112%	112%	107%	115%
	I bounce back after difficult times.	54%	97%	115%	117%	63%	113%	111%	112%	109%	109%
	I can achieve goals despite barriers.	61%	95%	115%	116%	58%	113%	112%	113%	107%	110%
	I come through difficult times with little trouble.	65%	104%	120%	120%	54%	113%	108%	113%	104%	99%
Resilience	I feel full of energy when I am at school.	74%	108%	125%	131%	1%	103%	116%	101%	137%	104%
	I feel bubbly and full of life at school.	54%	119%	129%	114%	9%	99%	120%	123%	138%	95%
	I feel wide awake when I am at school.	45%	127%	115%	103%	54%	112%	115%	105%	128%	95%
	I look forward to coming to school.	36%	107%	108%	97%	84%	99%	127%	95%	141%	106%
Learning goal	I try hard to do well when I am in class.	93%	106%	91%	112%	74%	110%	92%	112%	103%	106%
	I pay attention when I am in class.	93%	105%	96%	100%	82%	103%	94%	114%	101%	110%
	I listen carefully when I am in class.	90%	104%	97%	92%	87%	105%	98%	118%	101%	107%
Social harm	I have been verbally harassed.	113%	109%	79%	108%	55%	136%	76%	116%	73%	136%
	I have been physically harassed.	146%	103%	79%	117%	55%	124%	50%	133%	67%	126%
	I have been bullied online/social media.	161%	100%	93%	129%	62%	100%	63%	130%	35%	126%
	I have been picked on.	122%	122%	91%	141%	59%	118%	87%	116%	41%	103%
Disruptive behaviour	Used devices in class for non-education.	97%	124%	82%	102%	139%	86%	75%	121%	70%	104%
	I have been disrespectful to teachers.	137%	92%	75%	102%	154%	94%	67%	115%	68%	96%
	I have misbehaved in class.	122%	81%	78%	115%	152%	101%	73%	110%	75%	93%
	I have ignored the teachers' instructions.	137%	89%	81%	121%	128%	96%	72%	113%	73%	90%
Risky behaviour	I smoked cigarettes/e-cigarettes/vaped.	263%	90%	51%	79%	101%	85%	69%	119%	36%	108%
	I drank alcohol.	237%	61%	60%	72%	122%	89%	68%	138%	36%	117%
	I used drugs.	308%	58%	53%	83%	95%	93%	51%	112%	34%	113%
	I skipped lessons/did not go to school.	190%	92%	63%	95%	138%	84%	75%	109%	52%	100%
	I damaged school property on purpose.	224%	73%	61%	92%	138%	88%	59%	127%	47%	92%

The data reveals significant differences in attributes between each school's positive and negative clusters. For example, attributes related to self-efficacy, such as "At school, even if the work is hard, I can learn it", demonstrate substantial variation. School G registers 119% of the mean, indicating a pronounced divergence in self-efficacy between the two clusters. In contrast, School E presents a distance of 66%, suggesting a more homogeneous student population. Similar patterns emerge for other self-efficacy attributes, where School E consistently reports smaller centroid distances than schools such as School G, which exhibits much more significant differences.

Attributes connected to vigour also demonstrate marked variability across schools. School D records exceptionally high centroid distances, such as 140% for "I am determined to achieve my goals", reflecting intense polarisation in the students' determination and resilience. Conversely, School E exhibits lower distances, such as 59% for the same attribute, indicating narrower differences between the clusters. Resilience-related attributes further emphasise these school-specific differences. For example, School I reports some of the largest distances in the dataset, including 137% for "I feel full of energy when I am at school". At the same time, School E stands out with exceptionally low values, such as 1% for the same attribute, demonstrating an almost negligible distinction between the two clusters.

The attributes measuring social harm provide further insights into the variability between schools. For example, School F demonstrates pronounced differences, including 136% for "I have been verbally harassed", whereas School D shows a similarly high value of 141% for "I have been picked on". In contrast, School I shows relatively minor differences in these measures, such as 41% for "I have been picked on", suggesting a more uniform student experience. The data regarding risky behaviour reveals extreme polarisation in some schools. School A reports a distance of 263% for "I smoked cigarettes/e-cigarettes/vaped", which exceeds the distances reported by other schools, indicating a sharp divide between the clusters. Conversely, School I reflects one of the lowest distances for this attribute at 36%, suggesting a more consistent pattern of behaviour across clusters.

Variability in disruptive behaviour also reflects distinct patterns across schools. School E shows the highest centroid distances, such as 154% for "I have been disrespectful to teachers", indicating substantial differences in behaviour between the clusters. In contrast, School G reports relatively smaller distances for disruptive behaviours, such as 67% for the same attribute, highlighting greater homogeneity in student experiences.

These differences in centroid distances illustrate how much the positive and negative clusters diverge within each school. Schools with larger distances, such as School A in risky behaviour or School I in resilience, exhibit greater polarisation, indicating that distinct subpopulations of students have markedly different experiences or outcomes. In contrast, schools with smaller centroid distances, such as School E across multiple attributes, demonstrate greater uniformity between clusters. These patterns are critical for understanding how school-specific contexts influence student outcomes and for identifying areas that may affect the accuracy and performance of predictive models such as GBTs.

In addition to school-specific intra-cluster distances (or ranges), the percentage of students allocated to each cluster is also school specific. Figure 4.5 shows the relative proportion of students in the positive and negative clusters at each school. School A has the highest percentage of students in the positive cluster (77%), whereas School I has the lowest proportion of students in the positive cluster (48%).

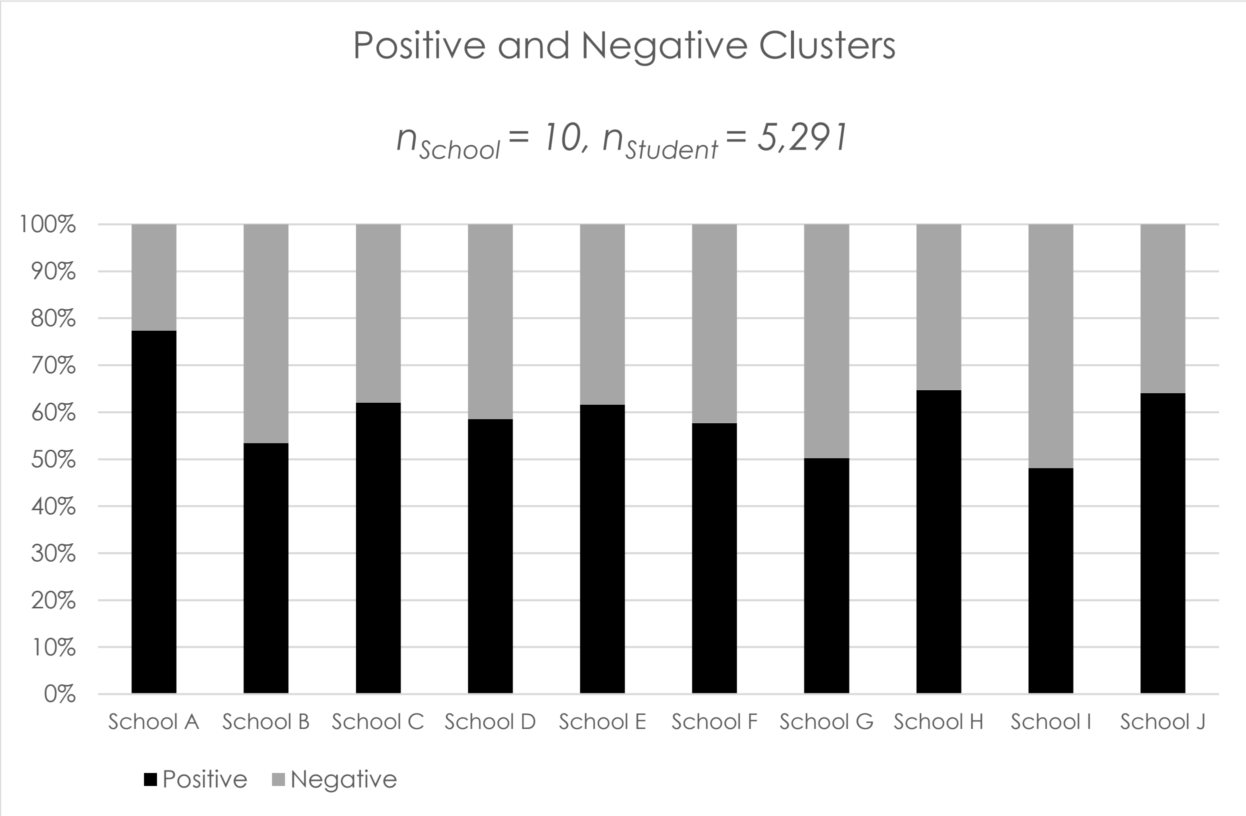

Figure 4.5. The relative proportion of students reporting positive and negative outcomes

School A has the most significant difference in the proportion of students allocated to positive and negative clusters as well as the largest centroidal distances between the clusters. That is, the student population relative to the other schools is less balanced, and the differences between the two clusters are most pronounced. Conversely, School I has the most symmetrical population and the greatest number of attributes where the two clusters report similar outcomes.

4.3 Predictive analysis

This section describes the GBTs' results in predicting the outcome cluster to which students are assigned, using the responses to the School Climate survey as the input. The next section details the GBTs' predictions.

GBTs were implemented using RapidMiner, a data mining and predictive analytics toolset capable of executing numerous machine learning models. GBTs were chosen because they can handle complex and dynamic relationships and have proven performance (compared to other classification techniques) in predicting student outcomes.

Table 4.2 presents the performance of the GBT models sorted in order of descending AUC. It also includes the number of predictions available for each school. This figure represents approximately 30% of the total responses, reflecting the division of the dataset into training and testing sets. As with the cluster distances examined in the previous section, the dataset size available for prediction significantly influences predictive performance.

Table 4.2. Results of the predictive analysis

School	Predictions	Sensitivity	Specificity	Precision	Accuracy	AUC
D	183	63.2%	84.1%	73.8%	75.4%	0.847
G	135	78.1%	76.1%	74.6%	77.0%	0.832
F	158	76.3%	69.2%	78.0%	73.4%	0.820
I	175	74.7%	76.2%	77.3%	75.4%	0.814
B	108	75.6%	71.6%	62.0%	73.1%	0.790
C	235	78.8%	64.0%	78.2%	73.2%	0.774
J	228	57.6%	72.8%	46.3%	68.4%	0.773
A	123	21.4%	93.7%	50.0%	77.2%	0.720
H	185	52.2%	74.6%	53.8%	66.5%	0.716
E	56	50.0%	66.7%	45.5%	60.7%	0.612
Whole-of-cohort	1,588	77.1%	64.9%	74.0%	71.8%	0.798

Table 4.2 offers a detailed evaluation of the GBT models trained to classify students into one of two outcome classes based on their experiences of School Climate. A separate model was developed for each school, followed by a final model trained on the whole-of-cohort. Metrics such as sensitivity, specificity, precision, accuracy, error, and the AUC enable a comprehensive assessment of the predictive quality of these models. By situating the AUC values within the AUC codex classifications, it is possible to evaluate their overall utility for predictive purposes. The codex categorises AUC scores as follows: 0.9–1.0 indicates “excellent” predictive quality, 0.8–0.9 “very good”, 0.7–0.8 “good”, 0.6–0.7 “satisfactory”, and 0.5–0.6 “unsatisfactory” (Trifonova, et al., 2013, p. 182).

Table 4.3. presents an assessment of GBT classification quality according to AUC value.

Table 4.3. Assessment of GBT classification quality according to AUC value

AUC values	Predictive quality	School
0.9–1.0	Excellent	-
0.8–0.9	Very good	D, G, F, and I
0.7–0.8	Good	B, C, A, J, and H
0.6–0.7	Satisfactory	E
0.5–0.6	Unsatisfactory	-

Sensitivity, which measures the models' ability to identify positive cases correctly, shows considerable variation across schools. School C achieves the highest sensitivity (78.8%), closely followed by Schools G (78.1%) and F (76.3%). These results highlight the ability of these models to effectively detect positive cases, which is critical for outcome prediction. Conversely, School A performs poorly on this measure (21.4%), indicating significant limitations in identifying positive cases. Specificity, which measures the ability to classify negative cases correctly, reveals an opposite trend for School A, which achieves the highest specificity (93.7%), indicating that it prioritises identifying negative cases at the expense of positive classification. School D, with a specificity of 84.1%, demonstrates a better balance between positive and negative case detection. Schools E and J exhibit comparatively weak specificity (66.7% and 72.8%, respectively), signalling challenges distinguishing between classes.

Precision, which captures the accuracy of positive predictions, provides additional insight into the models' utility. Schools F and C achieve the highest precision (78.0% and 78.2%, respectively), reflecting their reliability in making accurate positive predictions. By contrast, Schools E and J exhibit lower precision (45.5% and 46.3%, respectively), indicating higher false positive rates. These lower precision scores correspond to weaker AUC values, which, according to the scoring criteria in Table 4.3, suggest "satisfactory" or "unsatisfactory" performance. Accuracy, representing the proportion of correct predictions overall, reinforces these trends. Schools A, G, and D achieve the highest accuracy (77.2%, 77.0%, and 75.4%, respectively), reflecting their reliability in predicting both positive and negative outcomes. School E, with an accuracy of only 60.7%, performs poorly in this regard, which aligns with its low specificity and precision.

Error rates, the inverse of accuracy, further support this analysis. Schools A, G, and D exhibit the lowest error rates (22.8%, 23.0%, and 24.6%, respectively), confirming their overall predictive consistency. School E, however, has the highest error rate (39.3%), demonstrating its significant limitations as a classifier. According to the AUC codex, high error rates often correlate with models classified as "satisfactory" or "unsatisfactory", which is consistent with School E's low AUC of 0.612.

AUC values serve as the most comprehensive measure of model performance, summarising the balance between sensitivity and specificity and providing a robust indicator of overall predictive quality. School D achieves the highest AUC (0.847), categorising it as "very good" according to the codex. Schools G and F also perform strongly within this category (0.832 and 0.820, respectively), demonstrating consistently reliable predictive capacity. Schools A and C fall into the "good" range (0.720 and 0.774, respectively), reflecting acceptable but less robust discriminatory ability. By contrast, Schools E and H achieve the weakest AUC values (0.612 and 0.716, respectively), placing them in the "satisfactory" range. School E's performance is particularly concerning, as its AUC approaches the "unsatisfactory" threshold, indicating severe limitations in its ability to separate the two outcome classes effectively.

The model trained on the whole-of-cohort achieves an AUC of 0.798, positioning it in the upper "good" range. Its sensitivity (77.1%) and precision (74.0%) are strong, but its specificity (64.9%) lags behind that of the top-performing individual models, such as those for Schools D and A. These results suggest that the whole-of-cohort model generalises well across the dataset but sacrifices some local optimisation in exchange for broader applicability.

These results suggest that the GBT models provide substantial value for predicting student outcomes based on School Climate, particularly when tailored to individual schools. The models for Schools D, G, and F stand out as the strongest, achieving a "very good" predictive quality with balanced sensitivity, specificity, and precision. While achieving high specificity, School A balances this against weaker sensitivity, resulting in a "good" overall performance. Schools E and H, however, demonstrate significant challenges, with "satisfactory" AUC values reflecting broader limitations in their reliability. While useful as a generalised classifier, the model trained on the whole-of-cohort does not match the quality of the best individual school models.

In terms of their overall value, these GBT models demonstrate clear potential for predictive purposes, particularly when deployed in contexts that require a robust balance between sensitivity and specificity. High AUC values for the best-performing models (Schools D, G, and F) indicate strong predictive utility with reliable identification of positive and negative cases. The moderate performance of some schools (e.g. Schools A and C) further demonstrates the flexibility of these models in adapting to different data contexts, although variability in predictive quality remains. Models with lower AUC values, such as Schools E and H, highlight areas where improvement is necessary, whether through enhanced feature engineering, larger datasets, or adjustments to model parameters. As such, these models hold significant value as tools for outcome prediction, offering both specificity in local applications and broader generalisability where needed.

4.3.1 Gradient boosted trees: Results

The GBT modelling intends to predict the outcome cluster (positive or negative experience) to which each student belongs based on their individual responses to the School Climate survey. The reason for this is to understand the nature and strength of any relationship between school experience and the outcomes achieved (wellbeing and behaviour).

After completing the predictions, the model generates predictive weights for the attributes (variables). These predictive weights describe the linear relationships between each attribute and the frequency with which it supports or contradicts a prediction. However, this does not imply that the relationship between any given attribute and the outcome is inherently linear.

The GBT model was trained and tested separately for each school and for the whole-of-cohort. The analysis produced predictive weights for each question across the four School Climate pillars: Safety, Interpersonal Relationships, Teaching and Learning, and Institutional Environment. The following subsections provide a detailed breakdown of the predictive weights generated by the GBT model.

4.3.1.1 Whole-of-cohort

Figure 4.6 shows the predictive weights assigned by the GBT when predicting the outcome cluster to which student belongs. Each School Climate pillar is identified, as is the question to which the weight is assigned. The attributes with the most influence at the whole-of-cohort level are within the Institutional Environment pillar.

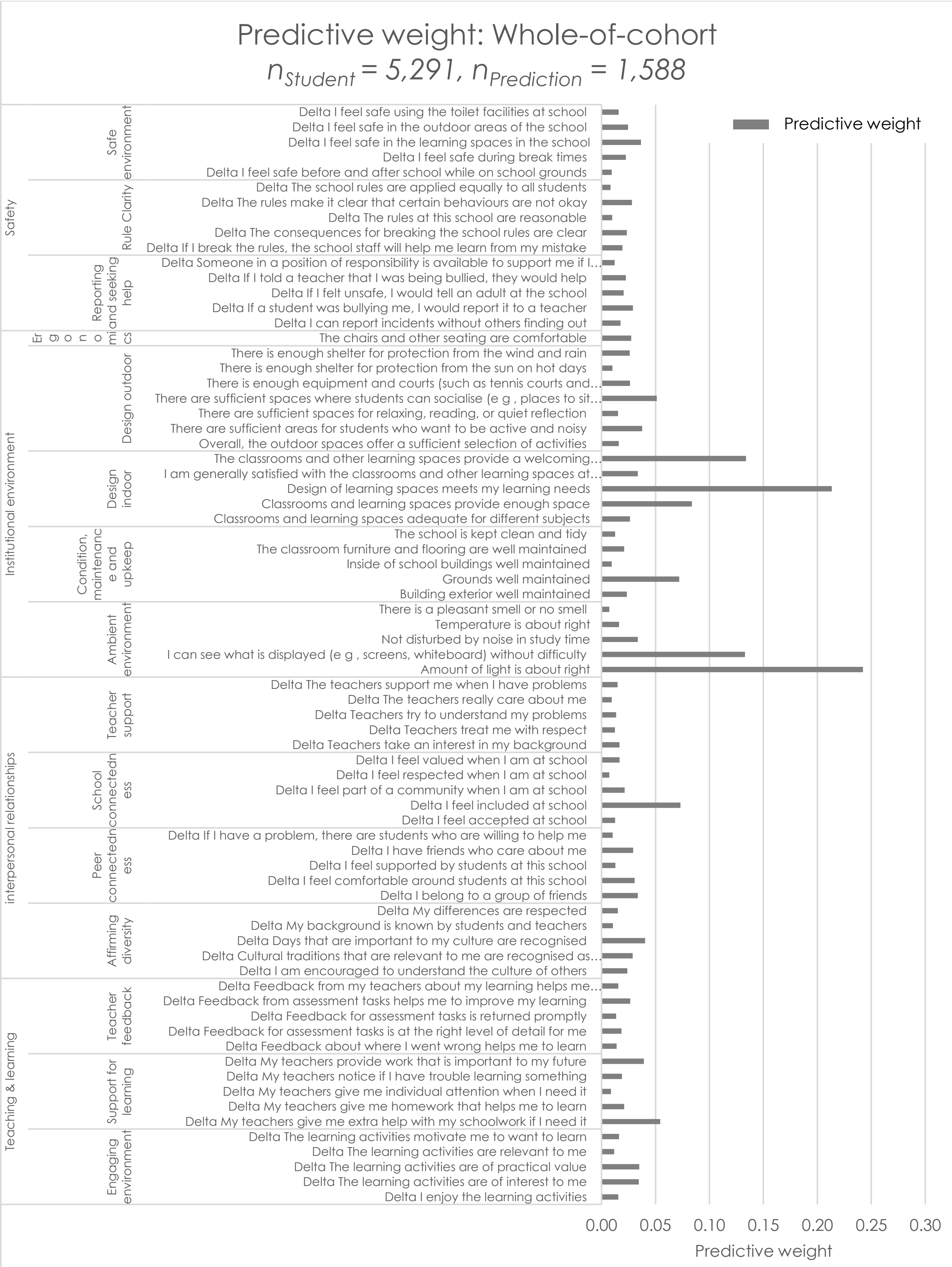

Figure 4.6. Whole-of-cohort predictive analysis

Table 4.4 lists (in descending weight order) the ten most heavily weighted attributes applicable to the whole-of-cohort, along with the associated pillar and construct. Also detailed is how much more heavily weighted (as a percentage) each attribute is compared to the next most heavily weighted attribute. The purpose of presenting this information is twofold. First, to show the relative difference in the importance of the most heavily weighted factors, with the four most heavily weighted factors being an order of magnitude more important than those following. Second, to show the different pillars represented by the most heavily weighted attributes. Seven of the ten most heavily weighted attributes are from the Institutional Environment pillar.

Table 4.4. Whole-of-cohort predictive analysis

Pillar	Construct	Attribute	Attribute weight	Percentage increase above the next attribute
Institutional Environment	Ambient environment	The amount of light is about right.	0.242	13%
Institutional Environment	Design indoor	Design of learning spaces meets my learning needs.	0.213	60%
Institutional Environment	Design indoor	The classrooms and other learning spaces provide a welcoming atmosphere.	0.134	1%
Institutional Environment	Ambient environment	I can see what is displayed (e.g. screens, whiteboard) without difficulty.	0.133	59%
Institutional Environment	Design indoor	Classrooms and learning spaces provide enough space.	0.084	14%
Interpersonal Relationships	School connectedness	Delta: I feel included at school.	0.073	2%
Institutional Environment	Condition, maintenance and upkeep	Grounds well maintained.	0.072	32%
Teaching and Learning	Support for learning	Delta: My teachers give me extra help with my schoolwork if I need it.	0.054	6%
Institutional Environment	Design outdoor	There are sufficient spaces where students can socialise (e.g. places to sit together).	0.051	27%
Interpersonal Relationships	Diversity	Delta: Days that are important to my culture are recognised.	0.040	3%

4.3.1.2 School-by-school

There are 73 attributes considered by the GBT models when predicting the outcome cluster to which each student belongs. Separate GBT models were developed for each school. Figure 4.7 shows how often each attribute is one of the ten most heavily weighted for a school. Institutional Environment attributes are presented in a darker shade of grey than those of the other School Climate pillars.

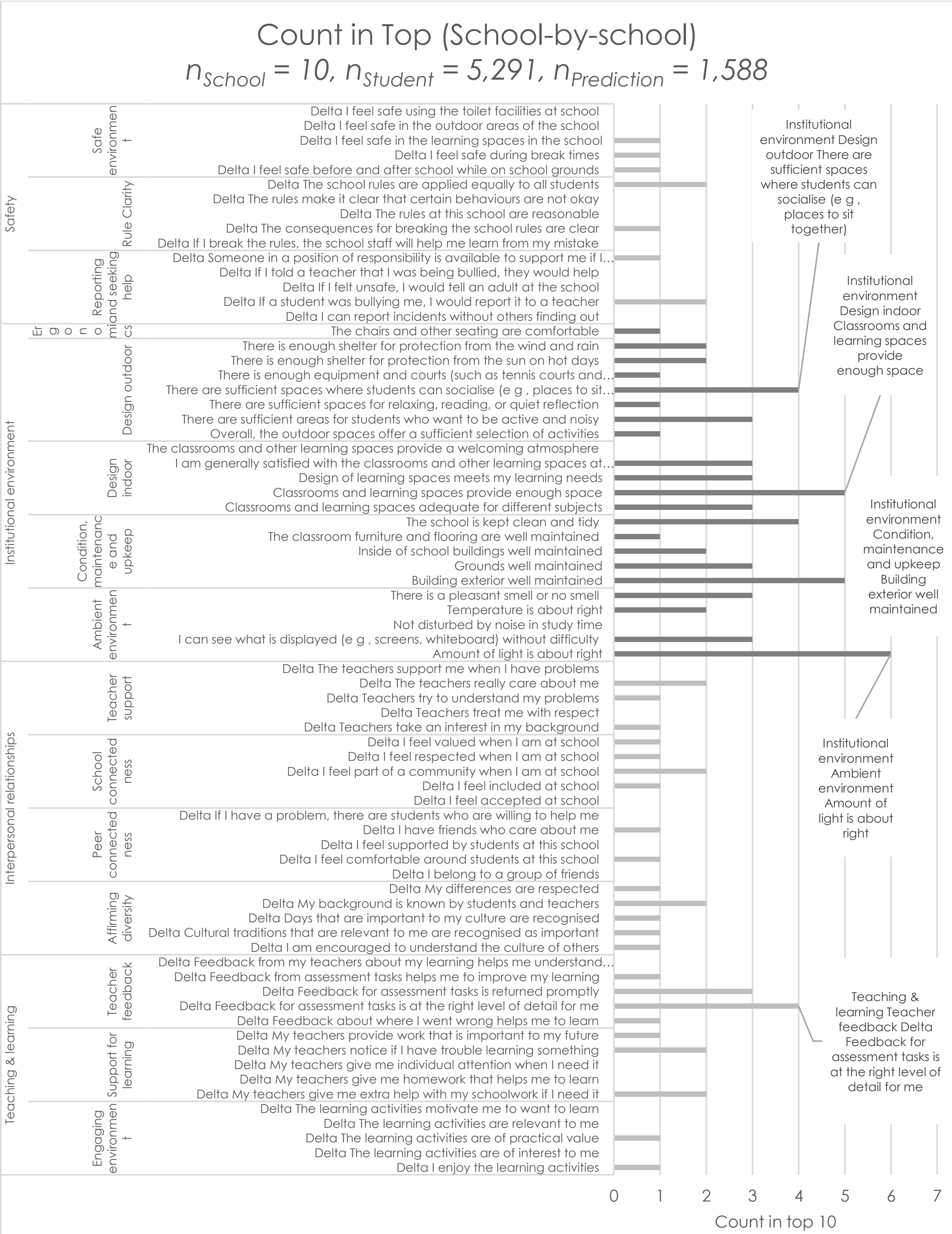

Figure 4.7. School-by-school predictive analysis

Attributes from the Institutional Environment pillar most frequently appear among the ten most heavily weighted attributes of the individual schools. Additionally, only Institutional Environment attributes rank among the ten most heavily weighted in more than four schools. The predictive weights for all School Climate attributes are shown in a heatmap included in Appendix 8.

Whereas Figure 4.7 shows all 73 School Climate attributes, Table 4.5 presents only Institutional Environment attributes. The rank order predictive weight for each Institutional Environment attribute is detailed for each school. Where an attribute is in the top ten highest weighted attributes, the cell is shaded grey.

Table 4.5. The predictive weight for each Institutional Environment attribute for each school

Construct	Attribute	School									
		A	B	C	D	E	F	G	H	I	J
Ambient environment	The amount of light is about right.	2	4	6	20	62	3	1	24	71	3
	I can see what is displayed (e.g. screens, whiteboard) without difficulty.	68	14	65	3	2	23	58	5	64	12
	Not disturbed by noise in study time.	16	43	11	50	32	20	60	65	56	55
	Temperature is about right.	13	66	52	33	15	5	62	1	19	72
	There is a pleasant smell or no smell.	8	1	63	63	31	70	70	45	13	10
Condition, maintenance and upkeep	Building exterior well maintained.	1	22	39	2	4	6	2	31	65	42
	Grounds well maintained.	29	34	15	5	72	65	56	7	8	28
	Inside of school buildings well maintained.	31	45	66	8	8	67	59	43	18	13
	The classroom furniture and flooring are well maintained.	4	58	26	65	40	32	64	15	39	56
	The school is kept clean and tidy.	40	49	37	4	5	17	66	16	4	2
Design indoor	Classrooms and learning spaces adequate for different subjects.	63	2	46	56	1	55	3	54	22	14
	Classrooms and learning spaces provide enough space.	44	15	13	6	66	19	4	2	1	9
	Design of learning spaces meets my learning needs.	61	39	1	71	52	52	55	53	3	4
	I am generally satisfied with the classrooms and other learning spaces at the school.	51	68	54	58	58	2	57	3	2	59
	The classrooms and other learning spaces provide a welcoming atmosphere.	52	32	16	67	50	61	65	56	47	22
Design outdoor	Overall, the outdoor spaces offer a sufficient selection of activities.	53	17	9	42	60	25	61	13	26	21
	There are sufficient areas for students who want to be active and noisy.	22	20	38	9	63	1	67	4	24	71
	There are sufficient spaces for relaxing, reading, or quiet reflection.	45	59	50	40	65	9	68	61	11	32
	There are sufficient spaces where students can socialise (e.g. places to sit together).	11	3	23	49	73	50	69	6	5	1
	There is enough equipment and courts (such as tennis courts and basketball hoops) for all who want to use it.	36	7	24	61	67	38	71	19	31	49
	There is enough shelter for protection from the sun on hot days.	26	29	17	1	7	37	72	55	29	45
	There is enough shelter for protection from the wind and rain.	9	62	71	29	9	56	73	57	43	46
Ergonomics	The chairs and other seating are comfortable.	34	60	36	36	14	48	63	70	6	40
Number of Institutional Environment attributes in the top ten		4	5	3	7	7	6	4	7	7	6

The results detailed in Table 4.5 reveal the relative importance of different infrastructure-related factors across the schools. While the specific rankings vary significantly, the results presented in the table demonstrate that Institutional Environment attributes consistently dominate the school's top ten most heavily weighted attributes, surpassing other the School Climate pillars of Safety, Interpersonal Relationships, and Teaching and Learning.

The focus on Institutional Environment attributes indicates their pivotal role in shaping the students' overall school experience. Key aspects such as the adequacy and maintenance of facilities, environmental comfort, and design features appear repeatedly in high-ranking positions, reflecting their importance. For example, the attribute "Amount of light is about right" is ranked as a top priority in multiple schools, securing first place for School G and appearing in the top ten for Schools A, B, F, and J. Similarly, "Building exterior well maintained" is highly rated, ranking first for School B and consistently appearing in the top ten for Schools D, E, F, and G.

However, the data also reveals significant variability in how schools prioritise specific attributes. For example, "Temperature is about right" ranks first for School H but falls as low as 72nd for School J. Similarly, sensory aspects such as "There is a pleasant smell or no smell" are ranked first by School B but receive much lower rankings in several other schools. These disparities suggest that while all schools recognise the importance of Institutional Environment attributes, their specific priorities reflect unique contexts and challenges.

The design of indoor and outdoor spaces is another area where variation is apparent. Attributes such as "Classrooms provide enough space" and "Design of learning spaces meets my learning needs" are highly ranked by Schools D, E, G, and I, underscoring the importance of functionality and adequacy in learning environments. Outdoor design attributes such as "There are sufficient areas for students who want to be active and noisy" and "There is enough shelter for protection from the sun on hot days" are also given prominence in certain schools, reflecting their role in supporting diverse student activities. However, the lower rankings these attributes receive in other schools highlight differences in local priorities, which may be influenced by climate, cultural expectations, or existing infrastructure.

Maintenance and upkeep emerge as another key theme. The attribute "The school is kept clean and tidy" achieves consistently high rankings, including second place for School J and top ten positions for Schools D, E, F, H, and I. This consistently high ranking suggests a universal recognition of cleanliness and upkeep as crucial components of a positive school environment. Similarly, "The classroom furniture and flooring are well maintained" is valued across several schools, demonstrating the importance of well-maintained physical spaces for comfort and satisfaction.

Despite differences in the rankings, the table indicates a clear trend: each school consistently prioritises multiple Institutional Environment attributes in its top ten rankings. Schools B, D, E, H, I, and J each feature seven Institutional Environment attributes in their top ten, while the other schools still include at least four or five. This repeated emphasis indicates that, regardless of variations in context, students at different schools consistently recognise the importance of their physical and sensory environments in shaping perceptions of school quality for both students and staff. At the same time, the variability in rankings demonstrates how each school tailors its priorities to address its unique circumstances and needs.

In summary, the table underscores the critical role Institutional Environment attributes play in defining School Climate and reveals how schools balance universal priorities with specific local considerations. The data highlights shared values and context-driven diversity, illustrating the complex interplay of factors that contribute to an effective and supportive school environment.

4.3.2 Complex and non-linear relationships

This research operates under the foundational assumption that the relationships between the independent variables (representing various aspects of School Climate) and the dependent variable (educational outcomes) exhibit complexity and non-linearity. This assumption profoundly influences the study's design and methodology in several key respects.

First, recognising the potential for non-linear relationships, the study employs GBTs as the primary predictive modelling technique. GBTs are particularly adept at identifying linear and non-linear relationships, as they do not presume a fixed form for the relationship among variables. Second, acknowledging complexity, particularly non-linearity, drives the decision to analyse each school individually. This individualised approach accounts for each school's unique contextual factors, ensuring that the specific dynamics influencing educational outcomes remain distinct and are not obscured by aggregating data across multiple institutions.

Table 4.5 details the rankings of predictive weights (Pre) and correlation coefficients (Cor) for each attribute of Institutional Environment, calculated separately for each school. These two measures provide distinct yet complementary insights into the relationships between variables. The predictive weight indicates the overall contribution of a variable to the model's predictive capacity, capturing both linear and non-linear patterns (represented by the hatched grey cells and plain grey cells in

Table 4.6, respectively). In contrast, the correlation coefficient quantifies the strength and direction of linear relationships. The analysis emphasises the key attributes by displaying in bold those that rank in the top ten for either predictive weight or correlation coefficient.

Table 4.6. Comparison pf predictive weights (Pre) and correlation coefficients (Cor) of School Climate attributes and educational outcomes

School		The amount of light is about right.	I can see what is displayed without difficulty.	Not disturbed by noise in study time.	Temperature is about right.	There is a pleasant smell or no smell.	Building exterior well maintained.	Grounds well maintained.	Inside of school buildings well maintained.	The classroom furniture and flooring are well maintained.	The school is kept clean and tidy.	Classrooms and learning spaces adequate for different subjects.	Classrooms and learning spaces provide enough space.	Design of learning spaces meets my learning needs.	I am satisfied with the classrooms and other learning spaces.	The classrooms and other learning spaces are welcoming.	Overall, the outdoor spaces offer a sufficient selection of	There are sufficient areas for students to be active and noisy.	There are sufficient spaces for relaxing, reading, or quiet	There are sufficient spaces where students can socialise.	There is enough equipment and courts.	There is enough shelter for protection from the sun on hot days.	There is enough shelter for protection from the wind and rain.	The chairs and other seating are comfortable.
A	Pre	18	64	22	1	7	12	24	11	2	49	5	58	36	67	20	14	43	32	66	23	31	21	68
	Cor	33	24	14	23	16	36	27	42	37	20	26	43	48	49	56	44	18	40	19	17	35	30	25
B	Pre	64	50	12	10	1	13	11	70	57	61	23	2	60	47	45	69	6	58	40	7	49	37	62
	Cor	28	25	6	5	41	32	14	48	49	21	15	38	34	39	54	36	19	26	9	12	30	31	16
C	Pre	21	5	59	14	65	23	58	62	8	49	64	42	6	1	32	2	15	19	7	28	73	26	46
	Cor	20	13	1	4	9	38	16	31	28	26	33	40	50	61	49	41	8	23	17	10	36	24	18
D	Pre	63	3	72	18	36	60	6	57	8	24	1	33	34	66	69	51	2	21	28	14	4	13	55
	Cor	23	19	5	16	11	26	20	37	29	21	24	40	44	42	49	32	9	31	17	12	18	8	7
E	Pre	40	2	25	62	31	71	9	6	61	64	37	48	57	65	58	26	1	13	7	56	3	24	15
	Cor	9	10	22	24	15	14	2	26	18	17	31	28	40	32	43	23	5	25	11	8	27	21	19
F	Pre	25	29	52	38	41	4	69	64	22	6	70	8	62	1	65	15	42	56	2	73	68	33	13
	Cor	38	11	4	9	14	32	22	45	39	20	31	40	46	48	56	49	34	23	28	6	15	5	24
G	Pre	29	56	52	63	3	5	45	32	42	7	14	10	4	38	2	62	57	41	1	27	39	30	66
	Cor	32	10	8	27	31	28	9	37	36	40	39	47	46	51	62	44	15	29	19	21	42	30	26
H	Pre	14	12	24	7	54	9	47	32	34	29	3	1	2	58	70	36	28	4	5	26	65	59	38
	Cor	13	8	9	12	24	20	15	28	31	23	25	41	45	39	49	42	6	18	17	21	37	40	26
I	Pre	52	15	58	18	44	12	4	31	32	2	6	14	1	13	10	67	5	42	3	24	34	53	45
	Cor	20	12	2	10	17	19	15	27	30	16	31	37	40	44	41	38	18	23	22	28	34	24	14
J	Pre	39	1	20	4	10	2	12	64	72	6	36	48	5	33	59	7	67	44	3	29	21	9	58
	Cor	8	4	2	3	6	13	1	19	22	10	12	34	30	39	35	31	11	29	15	7	16	14	23

The interaction between these two measures yields critical insights regarding the associations between variables and outcomes. Attributes with correlation coefficients ranking in the top ten are likely to support predominantly linear relationships with educational outcomes. Such correlation coefficients encapsulate linear patterns, where changes in one variable correspond to consistent, proportional changes in another variable. For example, a high positive correlation coefficient may suggest that improvements in a specific attribute of the school environment correspond with consistent enhancements in educational outcomes at a steady rate. An attribute that ranks in the top ten for correlation coefficient and predictive weight reinforces the notion that the linear aspect of the relationship primarily drives its predictive significance.

However, if an attribute ranks in the top ten for predictive weight but not for correlation coefficient, this implies that a non-linear relationship is more likely between the attribute and educational outcomes. The predictive weight computed by GBTs captures the variable's overall importance to the model's predictive capability, independent of whether the relationship is linear or non-linear. A high predictive weight alongside a low correlation coefficient indicates that the variable influences educational outcomes in a manner inadequately represented by a linear relationship. This relationship might manifest through thresholds, where an attribute begins to affect outcomes only beyond a certain level, or diminishing returns, where the effect lessens as the attribute increases. Alternatively, the relationship may involve interactions, where the impact of one attribute depends on the level of another. These patterns inherently highlight non-linearity, eluding capture by the correlation coefficient, which measures only direct, proportional changes.

This distinction underscores the significance of employing GBTs as the modelling approach. Traditional methods that rely exclusively on linear assumptions, such as multiple regression, fail to detect the intricate, non-linear patterns identifiable by GBTs. By integrating both predictive weights and correlation coefficients, this analysis presents a nuanced understanding of how attributes of the Institutional Environment contribute to educational outcomes.

4.4 Summary

This chapter presented a detailed analysis of the data collected from the School Climate survey, encompassing responses from ten government-funded secondary schools in NSW. The analysis began by assessing the surveyed schools and their representativeness compared to the broader NSW school population, followed by descriptive statistics of the School Climate experience.

Students were categorised using X-means clustering based on their self-reported wellbeing and behaviour outcomes. GBTs were then applied to predict students' outcome clusters based on their responses to the School Climate survey. The GBTs generated whole-of-cohort and school-specific predictive weights, indicating the strength of the predictive relationship between the School Climate attributes and the outcome clusters. The predictive weights emphasised the significant influence of the Institutional Environment pillar on student outcomes, particularly factors related to classroom design and maintenance. This chapter established a foundational understanding of the relationship between School Climate and student outcomes, which provides the basis for the subsequent discussion.

Chapter 5. Discussion

This chapter critically interprets the findings presented in Chapter 4, situating them within the broader theoretical frameworks and scholarly discourse introduced earlier in the thesis. The research examined the extent to which the students' experiences of school infrastructure, operationalised as the Institutional Environment within the School Climate framework, predict educational outcomes in secondary schools that are understood as complex and dynamic systems. The study responds to a growing recognition within educational research that schools are not uniform or mechanistic institutions (Bloom, 2019) but are contextually embedded systems in which outcomes emerge from multiple, interdependent variables (Rudasill, et al., 2018).

The interpretation of findings presented in this chapter is structured around four central insights that emerged from the analysis of the data. First, the clustering of student outcomes revealed two consistently distinct groups across all participating schools: students reporting predominantly positive wellbeing and behaviour outcomes, and students reporting negative outcomes. This distinct clustering reduced the dimensions, simplifying later analysis while also preserving the multidimensional information describing educational outcomes. Second, the predictive relationships between School Climate attributes and student outcomes were highly school specific and frequently non-linear, indicating that no single attribute exerted uniform influence across the different contexts. Third, infrastructure, captured through the Institutional Environment pillar of School Climate, consistently emerged as the strongest predictor of student outcomes. This finding challenges the prevailing assumptions about its peripheral role in school improvement. Finally, the methodological application of machine learning, particularly GBTs, demonstrated the analytical value of a modelling approach that is suited to analysing educational data without relying on a priori assumptions of linearity or generalisability.

The chapter starts by interpreting each of these findings and examining their significance in relation to existing theory and literature. It then articulates the study's contributions to knowledge across the *methodological*, *theoretical*, and *practical* domains. Two additional dimensions are introduced: a methodological reflection on the advantages and limitations of the analytical approach adopted and a brief discussion of the boundaries of interpretation and transferability of the findings. The chapter concludes with a synthesis that positions the findings within current debates regarding complexity, school effectiveness, and evidence-informed practice.

5.1 Interpreting the findings in context

The findings of this study contribute to a growing body of literature examining educational outcomes through the lens of complexity (van Vondel, et al., 2016; Koopmans, 2020). This analysis challenges the reliance on linear, generalisable models (Ansell & Geyer, 2017) by examining students' interactions with infrastructure and the influence these interactions have on students' wellbeing and behaviour. The findings demonstrate the importance of context-specific, non-linear dynamics in shaping student experiences and educational outcomes.

A defining feature of this research is its commitment to preserving the complexity inherent in educational outcomes. Rather than reducing these outcomes to isolated variables, such as test scores or attendance rates, the study employed cluster analysis to group students based on shared wellbeing and behaviour outcomes. X-means clustering enabled the construction of internally coherent outcome groups that retained the multidimensional nature of student outcomes while simplifying the prediction process. The analysis revealed two distinct clusters present in all ten schools: one characterised by more favourable self-reported outcomes and the other by less favourable responses. This demonstrates that the student population can be differentiated without resorting to arbitrary or reductive classifications. This approach aligns with a broader movement in the literature that calls for a shift towards more student-centred (Rudasill, et al., 2018), relational understanding of education, in which academic, social, and psychological dimensions are treated as interdependent (Bergman & Hupka-Brunner, 2013) and equally important.

The consistency of this outcome structure across the ten surveyed schools was notable; however, the composition of the clusters and the degree of separation between them was found to be school specific. This school specificity highlights the contextual embeddedness of student outcomes (Sellström & Bremberg, 2006) and supports the argument that schools operate as unique organisational ecosystems (Bloom, 2019). The observed differences suggest that while overarching patterns may exist across schools, their manifestation is shaped by local conditions (Steen, et al., 2013). These findings are consistent with the "school effect" noted by researchers examining the different outcomes achieved by different schools following system-wide policy interventions. These findings also resonate with complexity theorists such as Goldspink (2007) and Koopmans (2020), who emphasised the emergent nature of outcomes in educational systems and the importance of feedback loops, interactions, and threshold effects (Morrison, 2010).

Unlike other applications of cluster analysis in examining student outcomes (Wong, et al., 2022; Gonzalvez, et al., 2018; Dumuid, et al., 2017), the resultant clusters signified the beginning, not the end of the predictive process. The clusters provided the target for GBTs using School Climate attributes as inputs. Considering the school-specific clusters, the school-specific predictive relationships are to be expected. However, the predictive analysis also revealed the school-specific nature of the relationships (linear or non-linear) between School Climate attributes and student outcomes. Using GBTs, the study identified predictive relationships that varied in strength and nature across the schools. In most instances, the attributes found to be most predictive of outcomes in one school appeared insignificant in predicting outcomes in other schools. This variation challenges the assumption that changes to specific inputs, their value quantified through a universal effect size (Hattie, 2023), will predictably yield similar effects regardless of context. Furthermore, the comparison between correlation rankings and predictive weights supports the presence of non-linear effects, with several heavily weighted attributes exhibiting likely complex interactions, the importance of which is not readily ascertained using traditional regression techniques (Benkhalfallah & Laouar, 2023).

These findings provide empirical support for critiques of the EBPM approach as adopted in education, which most often relies on linear statistical methods to identify generalisable "what works" (Biesta, 2010) strategies. The inconsistent performance of strategies resulting from this approach (Fullan, 2019), particularly when transplanted across contexts, has been widely noted in the literature (Lingard, 2013; Loughland & Thompson, 2016; Morrison, 2021). This study supports the notion that educational outcomes are shaped by the complex interactions between attributes, which are unique to each school, rather than individual attributes working in isolation (Snyder, 2013). This interpretation supports the view that schools are complex, adaptive systems (Rudasill, et al., 2018) requiring contextually specific understanding and intervention (Steen, et al., 2013).

A significant finding emerging from the predictive analysis is the dominant predictive role of infrastructure, which was captured through the Institutional Environment pillar of School Climate. Across all schools, attributes related to the physical environment—such as lighting, cleanliness, spatial adequacy, and ambient comfort—emerged as the strongest predictors of student outcomes. This finding provides a counterpoint to the widely held view that the effect of infrastructure is below the threshold required for consideration as a potential policy lever (Hattie, 2023; Crampton, 2009; Hong & Zimmer, 2016). While prior studies have acknowledged that poor infrastructure may negatively affect learning (Uline & Tschannen-Moran, 2008; O'Neill & Oates, 2001), this study finds the experience of infrastructure to be a potentially significant predictor of student wellbeing and behaviour.

Moreover, the predictive importance of infrastructure was not uniform across the schools. While specific attributes, such as natural light and cleanliness, consistently attracted high predictive weights, others, such as temperature control and access to quiet spaces, varied in importance depending on the school. This school-specific variation further reinforces the centrality of context and highlights the limitations of generic or standardised approaches to infrastructure design and evaluation.

The predictive value of infrastructure experience as perceived by the school user is consistent with observations noting the importance of user satisfaction on attendance (O'Neill & Oates, 2001), behaviour (Berman, et al., 2018), and teacher turnover (Buckley, et al., 2005). These findings support the notion that performance improvement efforts targeting the physical environment benefit from being informed by user experience rather than relying solely on expert opinion (Roberts, 2009).

Machine learning methods, specifically GBTs, enabled the modelling of relationships between School Climate and educational outcomes without reliance on assumptions of linearity or uniformity (Yu & Ho, 2022). GBTs detect subtle patterns in the data, capturing the interaction effects and non-linear dynamics that typify educational systems (Yang, et al., 2020). While such models are not designed to establish causality (Delgado-Panadero, et al., 2022), they provide insight into which factors appear to be most influential within a given context. Applying them at the individual school level ensured that the models reflected each site's unique configurations rather than system-wide averages (Wu & Coggeshall, 2012). This approach represents a departure from the dominant practice of pooling data for cross-sectional meta-analysis and offers a more granular, locally attuned understanding of school dynamics (Farhood, et al., 2024).

In summary, the interpretation of the findings underscores the inadequacy of linear, one-size-fits-all models for explaining educational outcomes (Koopmans, 2020). Schools are complex systems where interdependent, context-sensitive factors shape student experiences and outcomes. The prominence of infrastructure, the variability of attribute–outcome relationships, and the observed non-linearity all point towards the need for appropriate frameworks and accompanying methodologies capable of accommodating complexity (Kariippanon, et al., 2018) rather than simplifying it.

5.2 Contributions

This study offers a multifaceted contribution to research exploring the contribution of infrastructure to educational outcomes, advancing theoretical discourse, refining methodological approaches, and generating insights for practice and policy. Integrating complexity theory (Koopmans & Stamovlasis, 2016), machine learning techniques (Larose & Larose, 2015), and School Climate constructs (Wang & Degol, 2016), the research addresses longstanding limitations in conceptualising, measuring, and predicting educational outcomes (Biesta, 2010). The contributions presented in this section are organised across three domains: *theoretical*, *methodological*, and *practical*. Each domain addresses specific gaps in the literature while reinforcing the broader claim that educational systems are dynamic, context sensitive, and inherently complex. By bridging conceptual innovation with empirical rigour, the study challenges dominant paradigms and proposes viable alternatives for how education can be studied, understood, and improved.

5.2.1 Theoretical contributions

The findings of this study contribute substantively to ongoing theoretical debates in educational research, particularly those concerned with the nature of school systems, the conceptual structure of School Climate, and the frameworks underpinning educational effectiveness. The study advances the theoretical discourse in three key areas: the operationalisation of complexity theory in empirical educational research; the integration of infrastructure into the School Climate framework; and the reorientation of outcome constructs towards holistic, student-centred models.

The findings of this study support the notion that schools are complex rather than complicated (Bloom, 2019). Complexity theory posits that outcomes are the emergent result of dynamic, context-dependent interactions between multiple interrelated elements in such systems (Mason, 2008). These systems are sensitive to initial conditions, exhibit feedback loops, and resist changes resulting from simplification through traditional causal models (Morrison, 2010). The school-specific and non-linear predictive relationships identified in this study provide strong empirical support for this perspective. Rather than demonstrating consistent, replicable associations between specific School Climate attributes and student outcomes, the analysis revealed considerable variation in the strength and nature of these relationships across the schools. This variation cannot be explained within the dominant EBPM paradigm (Fullan, 2019; Geyer, 2012), which assumes that generalisable causal relationships exist and can be codified into universally applicable interventions (Biesta, 2010).

The theoretical significance of these findings lies in their challenge to reductionist assumptions in educational effectiveness research. They affirm the argument, advanced by scholars such as Goldspink (2007), Johnson (2018), and Bloom (2019), that schools cannot be studied effectively using mechanistic or linear systems methodologies. Instead, the findings call for theoretical models that accommodate variability and interdependence—features that complexity theory is uniquely equipped to address (Koopmans, 2020). Notably, this study addresses a gap in the literature by advancing complexity theory beyond rhetorical invocation (Geyer, 2012), embedding it as a conceptual and explanatory framework for understanding educational outcomes.

A second theoretical implication relates to the positioning of infrastructure within the School Climate framework. The literature recognises four core dimensions of School Climate: Safety, Interpersonal Relationships, Teaching and Learning, and Institutional Environment (Wang & Degol, 2016). However, empirical studies have disproportionately focused on the first three (Zullig, et al., 2010; Thapa, et al., 2013), typically treating Institutional Environment (infrastructure) as a peripheral or modifying factor (Uline & Tschannen-Moran, 2008). This study challenges that implicit hierarchy. The analysis found that students' perceptions of their school's infrastructure consistently carried the most significant predictive weight regarding wellbeing and behaviour outcomes. The strength and consistency of this result, observed across ten diverse school settings, suggest that infrastructure should be treated not as an ancillary consideration but as a central pillar of School Climate. This approach has important implications for theoretical models of school experience. This study finds that physical space is not merely a backdrop to educational processes but an active component of the school ecology that interacts with social and pedagogical dimensions to shape student outcomes. The inclusion of infrastructure as a fully integrated component of School Climate theory advances the work of Wang and Degol (2016), who proposed its elevation to equal status but noted the absence of empirical studies demonstrating its centrality. This study complements the work of Aldridge et al. (2024), where Institutional Environment is considered but does not include infrastructure.

A third contribution to theory relates to the consideration and analytical treatment of educational outcomes. Much of the existing research focuses on academic achievement (Baldwin, et al., 2011; Brown, 2018), operationalised through standardised test scores (Chezan, et al., 2017; Sayer, 2020). While these metrics are thought to provide important insights, they are limited in scope and do not fully capture the breadth of what education seeks to achieve (Lingard, et al., 2013). Further, the utility and appropriateness of these metrics have been increasingly questioned (Bromley, et al., 2023; Chen & Luoh, 2009). This study contributes to an alternative tradition that views education as a multidimensional process encompassing cognitive, emotional, and behavioural development (Brière, et al., 2013; Daily, et al., 2019). By clustering students according to wellbeing and behaviour data patterns, the study reflects a shift towards student-centred (Zengaro & Warley, 2022) and holistic models of educational success.

Using student-reported wellbeing and behaviour as outcome constructs responds to longstanding critiques of reductive measurement practices. Scholars such as Anderson (2005), Hayward et al. (2015), and Zullig et al. (2010) have argued for broader definitions of success that incorporate elements of agency, mental health, social connectedness, and behavioural engagement. This study operationalises that argument by creating a multidimensional outcome variable that maintains the richness of the student experience while enabling predictive analysis. The theoretical contribution lies in validating an outcome model that reflects the lived student realities, reinforcing the move away from narrowly defined achievement metrics.

This study's theoretical implications support a reorientation of educational research towards frameworks that reflect the complexity of real-world school environments (Koopmans, 2020) and analytical techniques that are appropriate to this complexity (Zaman, et al., 2023). Numerous researchers have questioned the applicability of linear, generalisable models in supporting policy development in education and more generally (Geyer, et al., 2015; Ansell & Geyer, 2017; Koopmans, 2020). This study supports these findings and proposes alternative conceptions of inputs and treatment of outcomes. This approach finds that context is essential, and educational success is predicted by student experience, which is dominated by their perceived interactions with infrastructure.

5.2.2 Methodological contributions

This study makes a distinctive methodological contribution by aligning analytical strategies with the concept of schools as complex systems (Bloom, 2019; Koopmans, 2020). Rather than treating complexity as an abstract or rhetorical construct, the research embeds it directly within the design and execution of the analysis, offering an example of complexity operationalised in empirical educational research.

Machine learning techniques, specifically X-means clustering and GBTs, were selected for their capacity to model non-linear, context-dependent relationships without relying on prior assumptions regarding variable independence or linearity (Xu, et al., 2014; Gonzalvez, et al., 2018; Yang, 2024). These methods are particularly well suited to investigating educational environments where multiple, interdependent factors shape outcomes and where generalised models may fail to account for local variation (van Vondel, et al., 2016; Loughland & Thompson, 2016).

A key innovation lies in constructing a multidimensional, student-centred outcome variable using X-means clustering. Unlike traditional approaches that reduce outcome data to composite scores or unidimensional indices (Bergeron & Rivard, 2017; Putansu, 2020), this method preserved the diversity of the student-reported experiences related to wellbeing and behaviour. The clustering process also performed a type of dimensionality reduction while retaining interpretive value, identifying groups of students with shared outcome profiles based on the internal structure of the data (Wu & Coggeshall, 2012).

These clusters then served as outcome classes for GBT models, which were trained separately for each school. This school-specific modelling strategy represents a departure from pooled, system-wide analyses and enables the identification of patterns that may have remained obscured in aggregate models (Biesta, 2010; Waslander, et al., 2020). By capturing the distinctive relationships between School Climate variables and outcomes within each school, the analysis reinforced the theoretical framing of educational systems as contextually embedded and dynamically structured (Evers & Kneyber, 2015).

Alongside predictive accuracy, GBTs also offered interpretability through variable importance scores. These provided insight into the attributes most influential in shaping student outcomes in each context, supporting a more nuanced understanding of how different elements of School Climate interact in complex ways.

The methodological approach adopted in this study contributes to the expanding field of complexity-aligned educational research. It demonstrates how analytical strategies can be selected and configured to reflect the theoretical assumptions underpinning the study while also addressing the empirical realities of schooling. It offers an alternative to linear, generalisable models by emphasising analytical flexibility, contextual specificity, and methodological coherence.

5.2.3 Practical contributions

This study offers a set of practical contributions with direct implications for school improvement, educational policy, and the design of diagnostic tools. Most notably, it identifies infrastructure as the most powerful and consistent predictor of student outcomes, suggesting an urgent need to integrate physical environment considerations into strategic planning. In particular, the findings indicate that school improvement strategies may benefit from considering infrastructure as a core component of student experience rather than treating it as a logistical or aesthetic concern.

Furthermore, the school-specific nature of the relationships supports a move away from universalised intervention models (Geyer, 2012) towards school-led diagnosis (Steen, et al., 2013), using data that reflects context-specific dynamics (van Vondel, et al., 2016).

This research also illustrates the potential utility of predictive analytics in guiding school-level decision-making. The models developed in this study produced accurate predictions of student outcome clusters and identified the most influential attributes in each context. As such, these models provide a basis for developing a diagnostic approach using modified survey instruments that can assist school leaders in targeting interventions with greater precision. In schools and educational systems already engaged in regular School Climate data collection, the school-specific analytical approach adopted here could be readily integrated into existing practices, thereby enhancing the responsiveness and effectiveness of local decision-making.

5.2.4 Synthesising the contributions

Taken together, the contributions of this study represent a significant advancement in the conceptual, methodological, and practical understanding of educational systems, specifically in relation to the role of infrastructure and the utility of complexity-informed approaches.

Collectively, these contributions advance a nuanced, empirically grounded, and theoretically coherent understanding of educational improvement. By bridging the conceptual and the operational, this study presents a compelling case for reimagining how schools are studied and supported as complex systems whose effectiveness depends not on the uniform application of best practices, but on an attentive reading of context, the inclusion of student voice, and the careful alignment of theory, method, and practice.

5.3 Boundaries of interpretation

This section delineates the conceptual and epistemological boundaries within which the study's findings should be interpreted. It clarifies the extent to which the conclusions align with the study's theoretical positioning, highlighting the interpretive conditions that shape the meaning and scope of the results. Anchored in a pragmatic philosophical orientation and employing an inductive approach to theory development, this study advances interpretations premised on the view that knowledge arises through practical inquiry and is inherently situated within specific contexts of understanding.

Interpreting the findings of this study requires an understanding of the conceptual and epistemological boundaries that frame their meaning. These boundaries are not methodological limitations, but rather, the necessary interpretive conditions that define the scope and applicability of the results. Grounded in complexity theory (Mason, 2008) and context sensitivity (Koopmans, 2020) and guided by the pragmatic emphasis on actionable knowledge, the analysis positions educational outcomes as emergent, localised phenomena rather than generalisable effects.

The study employed a school-level modelling strategy that deliberately constructed predictive models for each participating school rather than aggregating data across contexts. This approach aligns with the theoretical commitment to situated complexity, privileging local dynamics over universal explanations. In line with inductive reasoning, patterns were allowed to emerge from the data rather than being imposed by preconceived hypotheses. As such, the findings should not be interpreted as indicative of general patterns across educational systems, but rather, as illustrations of how specific school environments shape student experiences. The school-specific nature of the results supports the central premise that educational phenomena are best understood in the context of immediate systemic conditions (Fullan, 2019).

A further interpretive boundary arises from the analytical emphasis on predictive association rather than causal explanation. The study employed GBTs to model the relationships between School Climate attributes and clustered student outcomes. While the models demonstrated high predictive performance, their purpose was diagnostic rather than explanatory. Variable predictive weights indicate which features contribute most to prediction accuracy, but they do not reveal causal mechanisms. This diagnostic orientation aligns with a pragmatic view that values predictive utility over definitive causal claims. The findings, therefore, provide insights into potential levers for intervention, but they should not be misinterpreted as establishing cause-and-effect relationships.

The outcome variable itself introduces an additional interpretive consideration. Rather than adopting traditional academic performance indicators such as grades or standardised test scores, the study constructed outcome clusters incorporating wellbeing and behaviour measures. This choice reflects a deliberate theoretical stance that challenges narrow metrics of success. Consistent with a pragmatic regard for outcomes that matter in lived experience, the study privileges broader conceptualisations of educational achievement. However, the interpretive relevance of these findings is limited to the domains aligned with this broader conceptualisation of educational outcomes.

Finally, the study's epistemological positioning places student perception at the centre of the analysis. Using School Climate data from student responses reflects a theoretical orientation that views learners as legitimate knowers of their educational experience. This positioning resonates with a pragmatic appreciation for stakeholder perspectives as valid sources of knowledge. As such, the findings speak to the students' lived realities within the participating schools. However, this approach intentionally foregrounds a specific perspective within the educational system and does not intend to represent a holistic institutional view.

By delineating these boundaries, the study clarifies the interpretive space within which its findings hold validity. Rather than seeking general laws or prescriptive solutions, the analysis demonstrates how a complexity-informed and pragmatically grounded approach can reveal the context-dependent nature of educational phenomena. Its inductive contribution lies in offering a model of inquiry that builds theory from situated observation, attending to local specificity and non-linearity.

5.4 Concluding the discussion

This chapter examined the study's key findings through theoretical, methodological, and empirical lenses, interpreting them within the broader scholarly discourse and reflecting on their scope and boundaries. Overall, the findings indicate that educational outcomes, defined as multidimensional clusters of wellbeing and behaviour indicators, are influenced by non-linear, school-specific relationships between School Climate attributes. Additionally, the students' experiences of infrastructure, referred to as the Institutional Environment, consistently appears as the most significant predictor in all contexts examined.

The study's contributions extend beyond the empirical confirmation of these relationships. Theoretically, it advances the argument that schools should be considered as complex adaptive systems in which outcomes are not the product of discrete causes, but rather, the result of interactions sensitive to context, time, and configuration. Methodologically, it demonstrates how machine learning techniques, specifically clustering and GBTs, can operationalise complexity theory in educational research, offering predictive insights while preserving student-level detail and contextual specificity. Practically, it identifies infrastructure as an often overlooked but highly consequential aspect of school experience and deserving of central attention in policy and school improvement efforts.

The analysis also draws attention to the epistemic boundaries of predictive modelling, the interpretive implications of perception-based data, and the limits of generalisability in systems defined by local complexity. The findings are not presented as prescriptive solutions but as evidence of what becomes visible when education is studied not through a reductive lens but as a dynamic and contextually embedded phenomenon.

The following chapter builds on this foundation, turning from interpretation to implications. It considers how the findings inform theory, policy, and practice and offers recommendations for future research. In doing so, it seeks not only to extend the significance of the current study but also to engage with the broader challenge of designing educational systems, policies, and interventions that are responsive to complexity, grounded in student experience, and informed by a richer conception of what matters in education.

Chapter 6. Conclusion

This final chapter consolidates the study's contributions and reflects on its broader implications for educational theory, policy, and practice. It builds on the context-specific, non-linear findings outlined in Chapter 4 and deepens the discussion of how school infrastructure and School Climate relate to student outcomes.

The chapter begins by exploring the theoretical, policy, and practical implications of modelling educational outcomes through a complexity-informed lens. It then critically examines the study's limitations, acknowledging the methodological and contextual boundaries that shape the findings. Key directions for future research are outlined, pointing towards areas for empirical expansion and conceptual refinement. Finally, the chapter synthesises the central insights of the thesis, explicitly addressing the research question and considering what this means for the future of educational systems that aspire to be evidence-informed and context-responsive.

6.1 Implications

The findings of this study have implications across multiple domains of education, including theory, policy, and practice. By uncovering the complex, context-specific nature of the relationships between School Climate attributes and student outcomes, the research challenges conventional models of educational effectiveness and calls for a paradigmatic shift in how improvement is conceptualised and operationalised. This section covers these implications, demonstrating how the empirical insights derived from this study can inform and reshape theoretical frameworks; guide more responsive policy decisions; and support school-level practices that are attuned to students' lived experiences. In doing so, it highlights the importance of engaging with complexity, prioritising local variation, and re-evaluating the role of infrastructure as a central determinant of student wellbeing and behaviour.

6.1.1 Implications for educational theory

The findings of this study carry important implications for how educational theory, particularly effectiveness and improvement, is conceptualised and developed. The study challenges the longstanding reliance on generalisable, input–output causation models by demonstrating that the relationship between School Climate attributes and educational outcomes is non-linear and school specific. These results call for a theoretical realignment that foregrounds complexity, acknowledges variability, and positions educational outcomes as emergent properties of dynamic systems rather than as predictable results of isolated inputs.

In the context of educational effectiveness research, the assumption that causal relationships can be extrapolated across schools has informed both theoretical frameworks and empirical methodologies. This study challenges that assumption. It provides empirical support for theories that conceptualise schools as complex adaptive systems. As such, the findings strengthen the case for adopting complexity theory as both a conceptual lens and a foundational principle in developing educational models. Future theoretical frameworks could benefit from moving beyond a reliance on reductionism, and instead account for interdependence and context-specific influence patterns.

The finding that infrastructure emerged as a strong and consistent predictor of student wellbeing and behaviour outcomes suggests the value of a more integrated theoretical consideration of physical space within School Climate models. While infrastructure has generally been treated as a contextual or background factor, the results of this study indicate that the students' experiences of the physical environment may play a more active role in shaping outcomes than previously recognised. As such, there is merit in exploring theoretical frameworks that explicitly incorporate infrastructure as a component of School Climate, acknowledging its potential interactions with interpersonal, pedagogical, and safety-related dimensions.

Finally, the successful application of machine learning methods in this study highlights the importance of aligning theoretical commitments with methodological choices. If educational systems are indeed complex and dynamic, as scholars increasingly suggest, it is important that theoretical frameworks accommodate analytical tools suited to modelling such complexity. This pivot to complexity does not imply a wholesale replacement of traditional methods, but rather, a recognition that theory and method will benefit from evolving in tandem. A complexity-informed theoretical stance invites models that are sensitive to non-linearity and responsive to contextual variation. In this way, theoretical advancement and methodological innovation are mutually reinforcing.

In summary, the findings of this study point towards the potential value of reorienting aspects of educational theory. Future conceptual frameworks may benefit from engaging more directly with complexity, not only as a descriptive feature of educational systems but also as a generative lens through which schools can be better understood. Such a shift would involve reconsidering some of the causal assumptions embedded in existing models, giving greater attention to underexplored dimensions such as infrastructure; broadening prevailing conceptions of student outcomes; and adopting analytical approaches that better reflect the dynamic, interconnected nature of educational settings.

6.1.2 Implications for educational policy

The findings of this study offer several important implications for educational policy, specifically how infrastructure is valued, how school improvement is approached and how systems account for local complexity. The consistent identification of infrastructure experience as the most powerful predictor of student outcomes across ten diverse schools necessitates a reconsideration of how infrastructure is positioned within policy frameworks. Traditionally treated as a background condition or a logistical necessity, infrastructure has generally been excluded from policies targeting school performance and educational equity. These findings challenge that marginal status and argue for its elevation as a policy priority.

The first implication relates to the allocation of resources. Infrastructure investment is typically framed in economic or compliance terms, with limited attention given to its educational impact. In systems shaped by evidence-based policy discourse, funding decisions are typically justified by referencing interventions that have perceived direct demonstrable effects on learning outcomes, usually academic achievement. However, this study demonstrates that students' infrastructure experiences are more predictive of wellbeing and behaviour than any other School Climate factor. As such, infrastructure emerges as a potential area for targeted investment, particularly in disadvantaged school contexts, where inadequate physical conditions can intensify other challenges and contribute to widening inequalities.

Secondly, the school-specific nature of the predictive relationships uncovered in this research highlights the limitations of universal, top–down policy frameworks. Current policy models often assume that improvement strategies—whether pedagogical, structural, or environmental—can be applied uniformly across diverse school settings. However, the variation in which School Climate attributes most strongly predicted outcomes from one school to another suggests that such assumptions may be potentially flawed. This does not preclude the existence of system-wide goals or standards. However, it does call for a more flexible infrastructure of support, one that recognises the heterogeneity of schools and the complexity of the systems in which they operate.

Thirdly, these findings suggest reconsidering how educational systems evaluate performance and define improvement. Accountability mechanisms typically rely on narrow metrics, such as standardised test scores or attendance rates, which do not capture the full scope of students' lived experiences. This study demonstrates that wellbeing and behaviour, increasingly recognised as essential to long-term educational and life outcomes, are strongly shaped by the school's physical environment. Policies that overlook such factors risk incentivising superficial improvement while neglecting the foundational conditions for student success. Incorporating measures of student-reported experience, including infrastructure-related dimensions, into school evaluation frameworks would enable more holistic and meaningful assessments of school quality.

Finally, the successful application of machine learning in this study indicates the potential for more extensive use of advanced analytics in policy development. Rather than relying exclusively on aggregate statistics or generalised insights from meta-analyses, policymakers might benefit from incorporating localised, high-resolution data to inform more nuanced and contextually responsive interventions. Predictive modelling, for example, could support the prioritisation of resources, enhance monitoring of system equity, and assist in the early identification of emerging risks. Nonetheless, the adoption of such approaches warrants careful consideration, with particular attention to issues of transparency, ethical governance, and democratic accountability.

In summary, this study encourages a policy shift in three key areas: the revaluation of infrastructure as a central component of educational success; the move towards school-specific, data-informed decision-making; and the broadening of performance frameworks to capture the complexity of students' experiences of school. These implications point to the need for policy systems that are evidence-informed, complexity-aware, student-centred, and contextually responsive.

6.1.3 Implications for school-level practice

The findings of this study provide important implications for practice at the school level, particularly for leaders, teachers, and school improvement teams seeking to enhance student wellbeing and behaviour outcomes. The finding that infrastructure, as experienced by the students, is a powerful predictor of outcomes across ten different schools suggests that successful school improvement strategies could extend beyond pedagogical and relational interventions. Focusing solely on teaching quality, behaviour management, or student support is unlikely to address the full range of factors influencing student outcomes. Instead, the physical environment should be considered a fundamental component of the educational experience, with demonstrable effects on how students feel, behave, and engage with school.

A key implication for practice is integrating infrastructure considerations into school improvement planning. This study indicates that decisions related to maintenance, refurbishment, and spatial reorganisation should be guided not only by asset management cycles or budget limitations but also by data on student experience. Schools would benefit from adopting structured processes to capture and respond to students' perceptions of the built environment, including lighting, cleanliness, temperature control, classroom space, and outdoor access. Such data could be gathered through regular School Climate surveys to inform prioritisation in site planning and capital works.

Moreover, the school-specific nature of the findings reinforces the importance of school-specific, evidence-informed decision-making. Each school in the study exhibited a unique configuration of predictive relationships between School Climate attributes and student outcomes. This school specificity suggests that "off-the-shelf" improvement strategies, whether based on best practice models or centralised policy directives, are unlikely to be universally effective. The implication is that school-level change efforts would benefit from an analysis of local conditions, student perceptions, and relational dynamics within the school community.

In conclusion, the implications for practice are twofold: to recognise infrastructure as a strategic priority for student wellbeing and behaviour, and to embrace school-specific diagnosis and intervention. These approaches align closely with the broader call of this study to engage with the complexity of educational environments and to position student experience at the centre of school improvement.

6.2 Addressing the research questions

Having outlined the broader implications of the findings, it is now appropriate to return explicitly to the research questions. These questions sought to explore the relationship between school infrastructure and educational outcomes; the potential value of complexity thinking to understanding this relationship; and the value of integrating infrastructure into models of school experience. The discussion below addresses each question in turn.

6.2.1 **RQ:** ***What is the relationship between school infrastructure and educational outcomes?***

The findings of this study suggest that the students' perceptions of school infrastructure, captured through the Institutional Environment pillar of School Climate, are meaningfully associated with key educational outcomes, particularly wellbeing and behaviour. Across all ten participating schools, physical environment features such as natural light, spatial adequacy, cleanliness, and ambient comfort consistently attracted relatively high predictive weights in the modelling. These results indicate that the students' lived experience of infrastructure may constitute a significant aspect of their broader engagement with school life, with potential consequences for their psychological safety, behaviour, and emotional wellbeing.

Importantly, the results indicate that the role of infrastructure is school specific. While interpersonal relationships, pedagogy, and safety have received greater attention in the literature, Institutional Environment emerged in this study as a consistently influential, if previously under-recognised, dimension. This finding challenges conventional assumptions that position infrastructure as a neutral or background variable in determining educational outcomes, instead suggesting that the physical environment may exert a more active influence on student experiences. Moreover, the analysis indicates that these relationships are non-linear, school specific, and contingent on local context. Infrastructure was a significant predictor of outcomes in each school studied, but its relative importance and interaction with other School Climate attributes varied across sites. Such variation aligns with a complexity-informed understanding of schools as dynamic systems, where outcomes emerge from the interdependence of multiple, contextually specific variables.

The findings indicate a relational and emergent character to the relationship between infrastructure and educational outcomes. Infrastructure, as experienced by students, appears to be a consequential and under-theorised aspect of the educational environment. Its influence is not readily generalisable but is instead shaped by local configurations, contextual affordances, and the dynamic interplay of material, social, and pedagogical conditions. In this view, infrastructure is not a passive backdrop to schooling but an active element within a complex system of influences on student development.

6.2.2 SRQ 1: How does situating infrastructure within the broader context of student experience inform the understanding of its role in shaping educational outcomes?

The findings of this study suggest that integrating infrastructure into conceptual and empirical models of school experience offers a more comprehensive account of the conditions that shape student outcomes. Traditionally, models of School Climate have foregrounded dimensions such as teacher–student relationships, discipline, and academic expectations, with the physical environment often relegated to the status of a contextual background. This study challenges that framing by suggesting that infrastructure should be understood as a core component of school experience, with measurable and consequential effects on student wellbeing and behaviour.

By incorporating infrastructure more fully into models of School Climate, it becomes possible to capture a broader and more accurate representation of the factors that shape student experience. This integration enables recognition of the affective and sensory dimensions of schooling—those aspects of the environment that influence how students feel, how safe they perceive themselves to be, and how they relate to the institution. Such considerations are particularly relevant considering growing interest in student wellbeing, mental health, and school belonging as critical outcomes.

Furthermore, treating infrastructure as an integral part of School Climate opens new avenues for school improvement and educational equity. It encourages schools and systems to recognise that the physical environment is not merely an operational or logistical concern, but a domain that can actively support or undermine educational aims. In disadvantaged contexts, infrastructure may either compound or alleviate other challenges, making its inclusion in analytical models and policy frameworks worthy of consideration.

The study also highlights the potential value of student-reported data in capturing the subjective experience of infrastructure. By focusing on the students' perspectives, models of school experience can become more responsive to the lived realities of those most affected by educational environments and more sensitive to variations that might otherwise remain obscured.

In summary, integrating infrastructure into School Climate models provides both a conceptual refinement and a practical imperative. It enhances explanatory power; supports understanding of school experience; and suggests new directions for evidence-informed, contextually responsive educational practice.

6.2.3 SRQ 2: To what extent does a complexity-informed approach help to assess the contribution of infrastructure to educational outcomes?

Complexity theory provides a valuable conceptual lens through which to understand the influence of infrastructure on educational outcomes. Traditional approaches to school effectiveness have typically relied on linear, input–output models that assume stable causal relationships across contexts. Such frameworks tend to treat infrastructure as a background or enabling condition rather than an active component of the school system. In contrast, complexity foregrounds interdependence, emergence, and contextual variation, which enables a more nuanced understanding of how infrastructure may contribute to student outcomes in different ways across different settings.

Within a complexity-informed framework, infrastructure is conceptualised not as a discrete input with a fixed effect but as a dynamic and adaptive system component. The interactions among physical space, social relationships, institutional norms, and pedagogical practices mediate the influence of infrastructure. This study's findings are consistent with this view. Although infrastructure-related attributes were strongly associated with wellbeing and behaviour outcomes, their significance varied markedly between schools, reflecting school-specific conditions and relationships.

Complexity thinking also supports a shift in methodological orientation, encouraging the use of analytical tools that are sensitive to non-linearity and local variation. This study's application of machine learning models, specifically X-means cluster analysis and GBTs, supports this approach, enabling the exploration of complex patterns without imposing assumptions of generalisability. This methodological alignment reinforces the theoretical stance that educational phenomena are emergent and multifactorial and that their investigation requires tools capable of accommodating such complexity.

In summary, complexity thinking enriches the understanding of infrastructure by positioning it as a co-constitutive element of school experience. It invites attention to the situated, relational, and emergent ways in which the physical environment influences student wellbeing and behaviour, and by doing so, it provides a conceptual foundation for moving beyond reductionist accounts of educational causality.

6.3 Limitations

This section outlines the study's methodological and empirical limitations. It identifies constraints relating to data, sampling, research design, and analytic scope. Several methodological limitations constrain the empirical scope of this study. While the design and analytical choices were aligned with the theoretical aims, certain constraints must be acknowledged to contextualise the robustness and generalisability of the findings.

The exclusive use of student self-reported data introduces limitations in data triangulation. While the adapted School Climate survey was subject to reliability and validity checks, the reliance on student perceptions restricts the analysis to a single stakeholder group. The study cannot provide a comprehensive systems-level view of School Climate or functioning without corroborating perspectives from teachers, school leaders, or external observers.

The sample also presents limitations in scope and representativeness. Data was collected from ten secondary schools within a single educational jurisdiction. Although the schools varied demographically, the sample was neither randomised nor nationally representative. Moreover, most of the schools served relatively socioeconomically disadvantaged communities, which limits the generalisability of the findings to other types of schools, such as primary settings, elite institutions, or those operating in different policy or funding contexts.

Temporal constraints further limit the study's conclusions. The research employed a cross-sectional design, capturing a single wave of data collection. Consequently, the analysis cannot address changes over time or the temporal sequencing of variables. The absence of longitudinal data precludes the identification of developmental trajectories or the examination of delayed effects between School Climate attributes and student outcomes.

While machine learning techniques such as GBTs demonstrate predictive performance, the interpretability of these models remains constrained. GBTs are inherently complex and often function as "black boxes", providing limited insight into the internal pathways through which inputs influence outputs. Although variable importance metrics were useful for identifying influential attributes, the models do not provider explanatory clarity for practitioners or policymakers seeking to understand specific causal dynamics.

Finally, the data collected and analysed did not extend to other influential domains such as family background, teacher quality, school leadership, or curriculum design. These elements are widely recognised as integral to student outcomes and may interact meaningfully with School Climate. Their exclusion reflects the restricted focus of the study but also highlights potential areas for future inquiry.

Acknowledging these limitations ensures a transparent account of the study's methodological boundaries and signals avenues for further research. Addressing these constraints in future studies may enhance the findings' explanatory power and applicability across broader educational contexts.

6.4 Future research directions

This study has contributed to a growing body of work that seeks to understand schools as complex systems and highlight the role of infrastructure in shaping student outcomes. At the same time, it has opened several pathways for future research, specifically methodological development, theoretical refinement, and applied inquiry. This study offers five key directions for future research.

First, future studies might benefit from incorporating longitudinal designs to explore how changes in School Climate, and infrastructure in particular, affect student outcomes over time. While this study captured a snapshot of student experience, a temporal approach would allow researchers to examine whether improvements in perceived infrastructure precede measurable gains in wellbeing, behaviour, or attainment. Such research could also assess the durability of intervention effects and the timing of impact across different dimensions of School Climate. A longitudinal design would also enable the exploration of feedback loops, which is an important feature of complex systems that remains underexamined in empirical educational research.

Second, there is scope to expand this research through a mixed-methods approach. While machine learning offers substantial analytical power, it does not provide explanatory depth regarding the mechanisms through which infrastructure influences outcomes. Integrating qualitative methods—such as interviews, focus groups, or ethnographic observation—could illuminate the subjective meanings that students and staff attach to their environments and explain why particular physical attributes matter in specific contexts. Such an approach could also clarify how perceptions of infrastructure intersect with other dimensions of School Climate such as safety or teaching quality.

Third, future research could explore the development and testing of interpretable machine learning models that balance predictive accuracy with transparency. Techniques such as SHAP, Local Interpretable Model-Agnostic Explanations (LIME), or rule-based decision trees could enable researchers to model complexity while providing straightforward explanations of model logic. This approach would support the translation of predictive insights into actionable strategies, particularly for practitioners and policymakers who require clarity to make informed decisions.

Fourth, there may be a benefit in extending the surveyed population to include a broader and more diverse range of schools. The current study focused on secondary schools in a single jurisdiction, many of which serve disadvantaged communities. Replicating the study across primary schools, different educational systems, or internationally would allow researchers to assess the consistency of the findings and test whether the predictive strength of infrastructure holds across various cultural and policy contexts. A larger and more diverse sample of schools would make it easier to identify and model different types of schools, as differences between them are likely to become clearer when examining a broader population.

Fifth, future studies could investigate how infrastructure interacts with other school-level variables such as pedagogical practices, leadership styles, or student support systems. For example, does the impact of infrastructure increase in schools with more student-centred teaching models? Which leadership styles best enhance improvements in the physical environment? Such questions contribute to a more integrated understanding of how multiple elements of the school ecosystem interact to produce student outcomes. Systems-based approaches, including structural equation modelling or ABM, could complement machine learning in exploring these multi-level dynamics.

Collectively, these directions point towards a research agenda that is both empirically grounded and theoretically ambitious. By combining robust analytics with richer contextual analysis, future studies can continue to advance the understanding of how educational environments function—and how they can be transformed in ways that are responsive to complexity, grounded in student experience, and informed by evidence.

6.5 Conclusion

This chapter considered the study's broader significance and its implications for theory, policy, and practice while acknowledging its limitations and suggesting directions for future research. Collectively, these reflections conclude the thesis and reaffirm its central contributions to educational research and school improvement.

This research explored how students' lived experiences of school infrastructure, including elements captured by the Institutional Environment pillar of School Climate, contribute to educational outcomes within complex and dynamic school environments. Through a novel combination of clustering and machine learning techniques, the research generated four key insights:

1. Student outcomes can be meaningfully grouped into multidimensional clusters.
2. Relationships between School Climate attributes and educational outcomes are school specific and frequently non-linear.
3. Infrastructure, as experienced by students, consistently emerges as the strongest predictor of student wellbeing and behaviour.
4. Machine learning approaches provide valuable tools for modelling complexity at the school level.

These insights challenge the assumptions embedded in linear, generalisable models of educational effectiveness and suggest that greater attention should be paid to context-sensitive, student-centred, and complexity-informed approaches. Theoretically, the study extends the School Climate framework by providing empirical support for elevating infrastructure from a peripheral concern to a central determinant of student experience. Methodologically, it demonstrates the value of integrating machine learning techniques in school-level analysis to reveal nuanced patterns of interdependence. Empirically, it highlights that the physical environment, although context-specific, significantly influences educational outcomes.

These contributions carry important implications. For researchers, they open new avenues for investigating educational complexity and refining theoretical models. For policymakers, they highlight the potential benefits of revaluing infrastructure and adopting decentralised, data-informed strategies for school improvement. For practitioners, they emphasise the importance of embedding students' perceptions, particularly of the built environment, into strategic planning and practice.

The study's limitations—its reliance on perception data, cross-sectional design, and specific sampling frame—establish functional parameters for future research. They highlight the need for further empirical work across diverse systems by employing longitudinal and mixed-methods approaches to test, refine, and extend the insights presented here.

Overall, the thesis suggests that complexity should be considered not as a barrier to analysis but as a defining feature of educational systems. In doing so, it offers a direction for research and reform that is more attentive to context, responsive to student experience, and committed to fostering more equitable, sustainable, and effective learning environments. Meaningful and lasting educational improvement is likely to be achieved not through applying generalised abstractions but through careful engagement with the specific, material, and relational realities of school life. This study contributes to that broader aim.

References

Aldridge, J. & Ala'I, K., 2013. Assessing students' views of school climate: Developing and validating the "What's Happening In This School?" (WHITS) questionnaire. *Improving Schools,* 16(1), pp. 47-66.

Aldridge, J. & Fraser, B., 2018. Teachers' perceptions of the organisational climate: A tool for promoting instructional improvement. *School Leadership and Management,* 38(3), pp. 323-344.

Aldridge, J. & Galos, S., 2018. Development and validation of an instrument to assess primary school students' perceptions of the learning environment. *Learning Environment Research,* 1(21), pp. 349-368.

Aldridge, J. M., Blackstock, M. J. & McLure, F. I., 2024. School climate: Using a person–environment fit perspective to inform school improvement. *Learning Environments Research,* Volume 27, pp. 411-430.

Aldridge, J. & McChesney, K., 2021. Parents' and caregivers' perceptions of the school climate: Development and validation of the "Parent and Caregiver Survey" (PaCS). *Learning Environments Research,* 1(24), pp. 23-41.

Al-Khatri, H., Al-Saadi, S., Al-Akhzami, F. & Khan, H., 2024. A comprehensive objective and subjective assessment survey of indoor environmental quality in higher education classrooms in a hot arid climate. *Building and Environment,* 98(3), pp. 297-307.

Allemang, B., Sitter, K. & Dimitropoulos, G., 2022. Pragmatism as a paradigm for patient-oriented research. *Health Expectations : An international journal of public participation in health care and health policy,* 25(1), pp. 38-47.

Anderson, S., 2005. The relationship between student psychological wellbeing, behaviour and educational outcomes: A lesson from the MindMatters Plus demonstration schools. *Australian Journal of Guidance and Counselling,* 15(2), pp. 235-240.

Ansell, S. & Geyer, R., 2017. 'Pragmatic complexity' a new foundation for moving beyond 'evidence-based policy making'?. *Policy Studies,* 38(2), pp. 149-167.

Arocupita, B. C., Cahuana, M. V., Hilachoque, V. H. & Bolanos, M. C., 2024. Optimizing hyperparameters in machine learning models for accurate fitness activity classification in school-aged children. *International Journal of Advanced Computer Science and Applications,* 15(8), pp. 962-972.

Australian Bureau of Statistics, 2017. *Socio-Economic Indexes for Areas.* [Online]
Available at: http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa
[Accessed 12 April 2018].

Australian Curriculum Assessment and Reporting Authority, 2020. *myschool.edu.au.* [Online]
Available at: https://www.myschool.edu.au/
[Accessed 15 10 2023].

Baima, N. R. & Paytas, T., 2020. Plato's pragmatism: Rethinking the relationship between ethics and epistemology. 1st edition ed. United Kingdom: Routledge.

Baker, R. S., Inventado, P. S. & Veletsianos, G., 2016. Educational data mining and learning analytics: Potentials and possibilities for online education. In: *Emergence and Innovation in Digital Learning.* Athabasca: Athabasca University Press, pp. 81-98.

Baldwin, C., Bensimon, E., Dowd, A. & Kleiman, L., 2011. Measuring student success. *New Directions for Community Colleges,* Issue 153, pp. 75-78.

Balestra, S. & Backes-Gellner, U., 2017. Heterogeneous returns to education over the wage distribution: Who profits the most?. *Labour Economics,* Volume 44, pp. 89-105.

Ball, S., 2012. Global education Inc: New policy networks and the neoliberal imaginary. 1st edition ed. Oxford, UK: Routledge.

Barrett, P., Zhang, Y., Moffat, J. & Kobbacy, K., 2013. A holistic, multi-level analysis identifying the impact of classroom design on pupils' learning. *Building and Environment,* Volume 59, pp. 678-689.

Belmonte, A. et al., 2019. School infrastructure spending and educational outcomes: Evidence from the 2012 earthquake in Northern Italy. *Economics of Education Review,* Volume 75, pp. 1-13.

Benbenishty, R., Astor, R. A., Roziner, I. & Wrabel, S. L., 2016. Testing the causal links between cchool climate, school violence, and school academic performance: A cross-lagged panel autoregressive model. *Educational Researcher,* 45(3), pp. 197-206.

Benkhalfallah, F. & Laouar, M., 2023. *Predicting student exam scores: Exploring the most effective regression technique.* Algiers, International Conference on Networking and Advanced Systems.

Bergeron, P. J. & Rivard, L., 2017. How to engage in pseudoscience with real data: A criticism of John Hattie's arguments in Visible Learning from the perspective of a statistician. *McGill Journal of Education,* 52(1), pp. 237-246.

Bergman, S. & Hupka-Brunner, M., 2013. The Interplay between educational achievement, occupational success, and well-being. *Social Indicators Research,* 111(1), pp. 75-96.

Berman, J. D. et al., 2018. School environmental conditions and links to academic performance and absenteeism in urban, mid-Atlantic public schools. *International Journal of Hygiene and Environmental Health,* Volume 221, pp. 800-808.

Biesta, G. J. J., 2010. Why 'what works' still won't work: From evidence-based education to value-based education. *Studies in Philosophy and Education,* Volume 29, pp. 491-503.

Blanchenay, P., Burns, T. & Koster, F., 2014. Shifting responsibilities - 20 years of education devolution in Sweden: A governing complex education systems case study. *OECD Education Working Papers*, 23 July, pp. 1-41.

Blass, P. E., 2020. Debunking Hattie: Evaluating the contribution of academic studies to policy development and implementation in Australia. *Journal of Education and Social Policy,* 7(4), pp. 91-97.

Bloom, J., 2019. A batesonian perspective on qualitative research and complex human systems. In: Koopmans & Stamovlasis, eds. *Complex dynamical systems in education: Concepts, methods and applications.* Cham: Springer International Publishing Switzerland, pp. 23-37.

Bonal, X. & Tarabini, A., 2013. The role of PISA in shaping hegemonic educational discourses, policies and practices: The case of Spain. *Research in Comparative and International Education,* 8(3), pp. 335-341.

Bonjour, D. et al., 2003. Returns to education: Evidence from U.K. twins. *The American Economic Review,* 93(5), pp. 1799-1812.

Breakspear, S., 2012. The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance, Paris: OECD.

Brière, F. N., Pascal, S., Dupéré, V. & Janosz, M., 2013. School environment and adolescent depressive symptoms: A multilevel longitudinal study. *Pediatrics,* Volume 131, pp. 702-708.

British Educational Research Association, 2017. *The BERA/SAGE handbook of educational research.* 1st edition ed. London: SAGE.

Bromley, P. et al., 2023. Global determinants of education reform: 1960 to 2017. *Sociology of Education,* 9(2), pp. 149-167.

Brown, C., 2018. Achieving evidence-informed policy and practice in education: EvidencED. 1st edition ed. Bingley, UK: Emerald.

Bryman, A. & Bell, E., 2015. *Business research methods.* 4th edition ed. Oxford, United Kingdom: Oxford University Press.

Buckley, J., Schneider, M. & Shang, Y., 2005. Fix it and they might stay: School facility quality and teacher retention in Washington, D.C.. *Teachers College Record,* 107(5), pp. 1107-1123.

Capra, F., 1996. *The web of life : A new synthesis of mind and matter.* 1st edition ed. London: HarperCollins.

Cartwright, N. & Hardie, J., 2012. *Evidence-Based Policy: A practical guide to doing it better.* 1st edition ed. Oxford, UK: Oxford University Press.

Castillo, N. & Wagner, D., 2014. Gold standard?: The use of randomized controlled trials for international educational policy. *Comparative Education Review,* 58(1), pp. 166-173.

Centre for Education Statistics and Evaluation, 2014. *School improvement frameworks: The evidence base,* Sydney: NSW Department of Education and Communities.

Centre for Education Statistics and Evaluation, 2014. *What works best: Evidence-based practices to help NSW improve student performance,* Sydney: NSW Department of Education and Communities.

Centre for Education Statistics and Evaluation, 2015. *School assets and student outcomes,* Sydney: NSW Department of Education.

Centre for Education Statistics and Evaluation, 2020. *Classroom management: Establishing and maintaining a positive learning environment,* Sydney: Department of Education.

Centre for Education Statistics and Evaluation, 2020. *What works best in practice,* Sydney: NSW Department of Education.

Centre for Education Statistics and Evaluation, 2020. *What works best: 2020 update,* Sydney: NSW Department of Education.

Centre for Education Statistics and Evaluation, 2024. *Schools and students: 2023 statistical bulletin,* Sydney: NSW Department of Education.

Chen, H. T., Morosanu, L. & Chen, V. H., 2024. Aligning evaluation approaches with context: Reductionism, system thinking, and pragmatic synthesis. *Asia Pacific Journal of Education,* 44(3), pp. 540-551.

Chen, S. S. & Luoh, M. C., 2009. Are mathematics and science test scores good indicators of labor-force quality?, Taipei: Springer Science + Business Media.

Chezan, L. C. et al., 2017. Measuring outcomes in schools. In: 1. edition, ed. *Handbook of Australian School Psychology.* Vancouver: Springer International Publishing, pp. 663-689.

Christensen, T. & Laegreid, P., 2022. *Taking stock: New Public Management (NPM) and post-NPM reforms - trends and challenges.* 1st edition ed. Cheltenham, UK.: Edward Elgar Publishing.

Cilliers, P. & Preiser, R., 2010. *Complexity, difference and identity: An ethical perspective.* 1st edition ed. Dordrecht: Springer.

Clark, H. H. & Chase, W. G., 1972. On the process of comparing sentences against pictures. *Cognitive Psychology,* 3(3), p. 472–517.

Crampton, F. E., 2009. Spending on school infrastructure: Does money matter?. *Journal of Educational Administration,* 47(3), pp. 305-322.

Creswell, J. W., 2009. Research design: Qualitative, quantitative, and mixed methods approaches. 3rd edition ed. Thousand Oaks, California: SAGE.

da Santos, C. et al., 2020. School academic climate and oral health (tooth loss) in adolescents. *Plos One,* Volume 13, pp. 1-11.

Daily, S. et al., 2019. School climate and academic achievement in middle and high school students. *The Journal of School Health,* 89(3), pp. 173-180.

Daniel, B. K. & Harland, T., 2017. *Higher education research methodology.* 1st edition ed. London: Routledge.

Darling-Hammond, L., 1994. Performance-based assessment and educational equity. *Harvard Educational Review,* 64(1), pp. 5-30.

Davis, B. & Sumara, D. J., 2006. *Complexity and education : Inquiries into learning, teaching, and research.* 1st edition ed. Mahwah, NJ: Lawrence Erlbaum Associates.

Delgado-Panadero, A., Hernández-Lorca, B., Teresa García-Ordás, M. & Alberto Benítez-Andrades, J., 2022. Implementing local-explainability in gradient boosting trees. *Information Sciences,* Volume 589, pp. 199-212.

Devereux, P. J. & Fan, W., 2011. Earnings returns to the British education expansion. *Economics of Education Review,* Volume 30, pp. 1153-1166.

Dhar, R., Nowlis, S. M. & Sherman, S. J., 1999. Comparison effects on preference construction. *Journal of Consumer Research,* 26(3), pp. 293-306.

Donnelly, R. A. & Abdel-Raouf, F. F., 2016. *Statistics.* 3rd edition ed. Indianapolis: Penguin Random House.

Duit, A., Galaz, V., Eckerberg, K. & Ebbesson, J., 2010. Governance, complexity, and resilience. *Global Environmental Change,* 20(3), pp. 363-368.

Dumuid, D. et al., 2017. Academic performance and lifestyle behaviours in Australian school children: A cluster analysis. *Health Education & Behavior,* 44(6), pp. 918-927.

Dwivedi, S., Kasliwal, P. & Soni, S., 2016. Comprehensive study of data analytics tools (RapidMiner, Weka, R tool, Knime). Indore, IEEE Xplore.

Evans, A. et al., 2019. The role of health and social factors in education outcome: A record-linked electronic birth cohort analysis. *PLOS ONE,* 14(8), pp. 1-19.

Everitt, B., Landau, S., Leese, M. & Stahl, D., 2011. *Cluster analysis.* 5th edition ed. Chichester: John Wiley & Sons Ltd.

Evers, J. & Kneyber, R., 2015. *Flip the system: Changing education from the ground up.* 1st edition ed. London: Routledge.

Faems, D., 2020. Moving forward quantitative research on innovation management: A call for an inductive turn on using and presenting quantitative research. *R & D Management,* 50(3), pp. 352-363.

Fan, J., Upadhye, S. & Worster, A., 2006. Understanding receiver operating characteristic (ROC) curves. *Journal of the Canadian Association of Emergency Physicians,* Issue 1, pp. 19-20.

Farhood, H., Joudah, I., Beheshti, A. & Muller, S., 2024. Evaluating and enhancing artificial intelligence models for predicting student learning outcomes. *Informatics,* 11(46), pp. 1-17.

Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recognition Letters,* Volume 27, pp. 861-874.

Feldhoff, T. O. G. B. A., Radisch, F. & Maag Merki, K., 2021. Why must everything be so complicated? Demands and challenges on methods for analyzing school improvement processes. In: *Concept and Design Developments in School Improvement Research.* Zurich: Springer International, pp. 9-25.

Forrester, J. W., 1961. *Industrial dynamics.* 1st edition ed. Cambridge: Pegasus Communications.

Foster, C., 2023. Methodological pragmatism in educational research: from qualitative-quantitative to exploratory-confirmatory distinctions. *International Journal of Research & Method in Education,* pp. 1-16.

Friedman, J., Hastie, T. & Tibshirani, R., 2000. Additive logistic regression: A statistical view of boosting. *The Annals of Statistics,* 28(2), pp. 337-407.

Fullan, M., 2006. The future of educational change: Systems thinkers in action. *Journal of Educational Change,* Volume 7, pp. 113-122.

Fullan, M., 2019. The elusive nature of whole system improvement in education. *Journal of Educational Change,* 17(4), pp. 539-544.

Fuller, B., 2022. Are teachers still the problem? An Analysis of the NSW Education 'What Works Best' documents. *Australian Journal of Teacher Education,* 47(8), pp. 87-99.

Futrell, M., 1989. Fourth-wave education reform: Are we ready?. *The Education Digest,* 55(3), pp. 3-6.

Gaertner, H., Wurster, S. & Pant, H. A., 2014. The effect of school inspections on school improvement. *School Effectiveness and School Improvement,* 25(4), pp. 489-508.

Gell-Mann, M., 1994. The quark and the jaguar: Adventures in the simple and the complex. 1st edition ed. New York: W. H. Freeman.

Geyer, R., 2012. Can complexity move UK policy beyond 'Evidence-Based Policy Making' and the 'Audit Culture'? Applying a 'Complexity Cascade' to education and health policy. *Political Studies,* 60(1), pp. 20-43.

Geyer, R., Cairney, P. & Astill, S., 2015. *Handbook on complexity and public policy.* 1st edition ed. Cheltenham: Edward Elgar Publishing.

Gillespie, A., Glaveanu, V. & de Saint Laurent, C., 2024. *Pragmatism and methodology: Doing research that matters with mixed methods.* 1st edition ed. Cambridge, U.K.: Cambridge University Press.

Gislason, N., 2009. Mapping school design: A qualitative study of the relations among facilities design, curriculum delivery, and school climate. *The Journal of Environmental Education,* 40(4), pp. 17-34.

Glouberman, S. et al., 2016. Complicated and complex systems: What would successful reform of Medicare look like?. In: *Changing Health Care in Canada.* Toronto: University of Toronto Press, pp. 21-53.

Goldkuhl, G., 2012. Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems,* 21(2), pp. 135-146.

Goldspink, C., 2007. Transforming education: Evidential support for a complex systems approach. *Emergence: Complexity and Organisation,* 9(1), p. 77.

Gonzalvez, C. et al., 2018. A cluster analysis of school refusal behavior: Identification of profiles and risk for school anxiety. *International Journal of Educational Research,* Volume 90, pp. 43-49.

Gorard, S., 2013. Research design: Creating robust approaches for the social sciences. 1st edition ed. London: SAGE.

Gorard, S., 2020. Getting evidence into education: Evaluating the routes to policy and practice. 1st edition ed. New York, NY: Routledge.

Goss, A., 2024. "This space is definitely a difference": The effects of a new building on school climate. *Learning Environments Research,* 27(3), pp. 777-796.

Gottfredson, C. D. & Payne, C. N., 2005. School climate predictors of school disorder: Results from a national study of delinquency prevention in chools. *Journal of Research in Crime and Delinquency,* 42(4), pp. 412-444.

Government Schools Finance Statistics Group, 2021. *National Schools Statistics Collection (NSSC) Finance,* Canberra: Unpublished.

Grace, B., Gregory, T., Collier, L. & Brinkman, S., 2022. Clustering of wellbeing, engagement and academic outcomes in Australian primary schools. *Child Indicators Research,* Volume 15, pp. 2171-2195.

Gray, J. & Beresford, Q., 2008. A 'Formidable Challenge': Australia's quest for equity in Indigenous education. *Australian Journal of Education,* 52(2), pp. 197-223.

Groves, R. et al., 2014. *Survey methodology.* 1st edition ed. New Jersey: John Wiley & Sons.

Guarido, M., 2018. Machine learning in geoscience: Facies classification with features engineering, clustering, and gradient boosting trees, s.l.: Crewes.

Gunter, T. & Shao, J., 2016. Synthesizing the effect of building condition quality on academic performance. *Education Finance and Policy,* 11(1), pp. 97-123.

Gutacker, N., Kinge, J. M. & Olsen, J. A., 2023. Inequality in quality-adjusted life expectancy by educational attainment in Norway: An observational study. *BMC Public Health,* 23(1), pp. 805-814.

Gu, X. & Blackmore, K. L., 2015. A systematic review of agent-based modelling and simulation applications in the higher education domain. *Higher Education Research and Development,* 34(5), pp. 883-898.

Halpin, A. W. & Croft, D. B., 1963. *The organizational climate of schools,* Chicago: University of Chicago.

Hamandan, A. & Ahmad Ganai, N., 2023. Artificial intelligence algorithm comparison and ranking for weight prediction in sheep. *Nature,* Volume 13, pp. 1-13.

Hanuliaková, J. & Barnová, S., 2015. Acta educationis generalis. *Acta Educationis Generalis,* 5(1), pp. 68-73.

Hardy, I., 2013. A logic of appropriation: Enacting national testing (NAPLAN) in Australia. *Journal of Education Policy,* 29(1), pp. 1-18.

Hart, M., Moore, M. & Laverty, M., 2017. Improving Indigenous health through education. *Medical Journal of Australia,* 207(1), pp. 12-13.

Hascher, T., 2008. Quantitative and qualitative research approaches to assess student well-being. *International Journal of Education Research,* Volume 47, pp. 84-96.

Hattie, J., 2013. What is the nature of evidence that makes a difference to learning?. *Form@re - Open Journal per La Formazione in Rete,* 13(2), pp. 6-21.

Hattie, J., 2023. Visible learning: The sequel a synthesis of over 2,100 meta- analyses relating to achievement. 1st edition ed. New York: Routledge.

Hattie, J. A. C., 2009. Visible learning: A synthesis of 800+ meta-analyses on achievement. 1st edition ed. Oxford: Routledge.

Hayward, M., Hummer, R. & Sasson, I., 2015. Trends and group differences in the association between educational attainment and U.S. adult mortality: Implications for understanding education's causal influence. *Social Science & Medicine,* 127(1), pp. 8-18.

Head, B. W., 2016. Toward more "Evidence-Informed" Policy Making?. *Public Administration Review,* 76(3), pp. 472-484.

Heckman, J., Humphries, J. & Veramendi, G., 2018. Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy,* 26(S1), pp. 197-246.

Herrim, J., 2014. Why do East Asian children perform so well in PISA? An investigation of Western-born children of East Asian descent, London: University of London.

Heschong, L., Wright, R. L. & Okura, S., 2002. Daylighting impacts on human performance in school. *Journal of the Illuminating Engineering Society,* 31(2), pp. 101-114.

Holland, J. H., 1995. *Hidden order: How adaptation builds complexity.* 1st edition ed. New York: Helix Books.

Holmund, H., Mcnally, S. & Viarengo, M., 2010. Does money matter for schools?. *Economics of Education Review,* Volume 29, pp. 1154-1164.

Hong, K. & Zimmer, R., 2016. Does investing in school capital infrastructure improve student achievement?. *Economics of Education Review,* 53(1), pp. 143-158.

Hopson, L. M. & Lee, E., 2011. Mitigating the effect of family poverty on academic and behavioral outcomes: The role of school climate in middle and high school. *Children and Youth Services Review,* Volume 33, pp. 2221-2229.

Hoy, W., Hannum, J. & Tschannen-Moran, M., 1998. Organisational climate and student achievement: A parsimonious and longitudinal view. *Journal of School Leadership,* Volume 8, pp. 336-359.

Hulland, J., Baumgartner, H. & Smith, K. M., 2018. Marketing survey research best practices: Evidence and recommendations from a review of JAMS articles. *Journal of the Academy of Marketing Science,* 46(1), pp. 92-108.

Hurley, P., 2011. *A concise Introduction to Logic.* 11th edition ed. Boston: Wadsworth.

Ibrahim, B., Ahenkorah, I. & Ewusi, A., 2022. Explainable risk assessment of rockbolts' failure in underground coal mines based on categorical gradient boosting and SHapley Additive exPlanations (SHAP). *Sustainability,* Volume 14, pp. 1-16.

Institute of Public Affairs, 2023. *Latest PISA Rankings.* [Online]
Available at: https://ipa.org.au/publications-ipa/media-releases/latest-pisa-rankings
[Accessed 9 December 2023].

Jørgensen, J. V., 2024. The differential impact of evidence types in policy decisions: Comparing evidence priorities in employment and education ministries. *Policy Studies,* pp. 1-28.

Jensen, B. R. J. & Kemp, A., 2011. The real issue in school funding: An analysis of increasing government school expenditure and declining performance. *Australian Economic Review,* 44(1), pp. 321-329.

Jiang, X. et al., 2022. Teacher–student relationships and adolescents' school satisfaction: Behavioural engagement as a mechanism of change. *British Journal of Educational Psychology,* 92(4), pp. 1444-1457.

Johnson, E. S., 2008. Ecological systems and complexity theory: Toward an alternative model of accountability in education. *Complicity: An International Journal of Complexity and Education,* 5(1), pp. 1-10.

Johnson, R. B. & Onwuegbuzie, A. J., 2004. Mixed methods research: A research paradigm whose time has come. *Educational Researcher,* Volume 33, pp. 14-26.

Johnston, J., 2018. Australian NAPLAN testing: In what ways is this a 'wicked' problem?. *Improving Schools,* 20(1), pp. 18-34.

Just, M. A. & Carpenter, P. A., 1971. Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior,* 10(3), pp. 244-253.

Kapoor, N. R. et al., 2021. A review on indoor environmental quality of Indian school classrooms. *Sustainability,* Volume 12, pp. 1-43.

Kariippanon, K. E., Cliff, D. P., Lancaster, S. L. O. A. D. & Parrish, A. M., 2018. Perceived interplay between flexible learning spaces and teaching, learning and student wellbeing. *Learning Environments Research,* Volume 21, pp. 301-320.

Kazakov, R., Petrova, P. & Kazakova, Y., 2024. Understanding and managing socioeconomic systems behaviour: Applications of qualitative and quantitative system dynamics and agent-based modelling in healthcare and pharmaceuticals, finance, arts and culture, sociology and education systems. 1st edition ed. Cham: Springer International Publishing AG.

Klenowski, V. & Wyatt-Smith, C., 2012. The impact of high stakes testing: the Australian story. *Assessment in Education: Principles, Policy & Practice,* 19(1), pp. 65-79.

Knudsen, H., 2017. John Hattie: I'm a statistician, I'm not a theoretician. *Nordic Journal of Studies in Educational Policy,* 3(3), pp. 253-261.

Konold, T., Cornell, D., Jia, Y. & Malone, M., 2018. School climate, student engagement, and academic achievement: A latent variable, multilevel multi-informant examination. *AERA Open,* 4(4), pp. 1-10.

Koopmans, M., 2020. Education is a complex dynamical system: Challenges for research. *Journal of Experimental Education,* 88(3), pp. 358-374.

Koopmans, M. & Stamovlasis, D., 2016. Education as a complex dynamical system. In: Koopmans & Stamovlasis, eds. *Complex dynamical systems in education: Concepts, methods and applications.* Cham, Switzerland: Springer International Publishing AG, pp. 1-7.

Koster, A. et al., 2017. The role of agent-based simulation in education. *Communications in Computer and Information Science,* Volume 677, pp. 156-167.

Kotu, V. & Deshpande, B., 2015. Predictive analytics and data mining: Concepts and practice with Rapidminer. 1st edition ed. Waltham: Morgan Kaufman.

Kovács, L. & Ghous, H., 2020. Efficiency comparison of Python and RapidMiner. *Multidiszcip Tudományok (Multidisciplinary Sciences),* Volume 10, pp. 212-220.

Kulick, J., Prieger, J. & Kleiman, M., 2016. Unintended consequences of cigarette prohibition, regulation, and taxation. *International Journal of Law, Crime and Justice,* 46(9), pp. 69-85.

Kutsyuruba, B., Klinger, D. A. & Hussain, A., 2015. Relationships among school climate, school safety, and student achievement and well-being: A review of the literature. *Review of Education,* 3(2), pp. 103-135.

Lamprianou, I., 2012. Unintended consequences of forced policy-making in high stakes examinations: The case of the Republic of Cyprus. *Assessment in Education: Principles, Policy and Practice,* 19(1), pp. 27-44.

Larose, D. & Larose, C., 2015. *Data mining and predictive analytics.* 2nd edition ed. Hoboken: John Wiley and Sons.

Lehrl, S., Evangelou, M. & Sammons, P., 2020. The home learning environment and its role in shaping children's educational development. *School Effectiveness and School Improvement,* 31(1), pp. 1-6.

Lingard, B., 2013. The impact of research on education policy in an era of evidence-based policy. *Critical Studies in Education,* 54(2), pp. 113-131.

Lingard, B., Martino, W. & Rezai-Rashti, G., 2013. Testing regimes, accountabilities and education policy: Commensurate global and national developments. *Journal of Education Policy,* 28(5), pp. 1-18.

Lingard, B. & Rizvi, F., 2009. *Globalizing education policy.* 1st edition ed. Oxford: Taylor and Francis Group.

Lingard, B. & Sellar, S., 2013. 'Catalyst data': Perverse systemic effects of audit and accountability in Australian schooling. *Journal of Education Policy,* 28(5), pp. 634-656.

Lin, L. & Chu, H., 2018. Quantifying publication bias in meta-analysis. *Biometrics,* 74(3), pp. 785-794.

Liu, S. et al., 2018. Investigating the diffusion of agent-based modelling and system dynamics modelling in population health and healthcare research. *Systems Research and Behavioral Science,* 35(2), pp. 203-215.

Lobascher, S., 2011. What are the potential impacts of high-stakes testing on literacy education in Australia?. *Literacy Learning: The Middle Years,* 9(2), pp. 9-19.

Locke, E. A. & Latham, G. P., 2020. Building a theory by induction: The example of goal setting theory. *Organizational Psychology Review,* 10(3), pp. 223-239.

Locke, K., 2015. Pragmatic reflections on a conversation about grounded theory in management and organization studies. *Organisational Research Methods,* 18(4), pp. 612-619.

Loughland, T. & Thompson, G., 2016. The problem of simplification: Think-tanks, recipes, equity and 'Turning around low-performing schools.'. *The Australian Educational Researcher,* Volume 43, pp. 111-129.

Lyu, Y., 2024. Shifting research paradigms in educational research: From positivism to interpretivism. *Science, Technology and Social Development Proceedings Series,* Volume 2, pp. 11-14.

Manning, D., Raghavan, P. & Schutze, H., 2008. *Introduction to information retrieval.* 1st edition ed. Cambridge: Cambridge University Press.

Marchant, G. J., 2004. What is at stake with high stakes testing? A discussion of issues and research. *Ohio Journal of Science,* 104(2), pp. 2-7.

Marfo, P. & Okyere, G., 2019. The accuracy of effect-size estimates under normals and contaminated normals in meta-analysis. *Helyon,* 5(6), pp. 1-9.

Martins, M. et al., 2021. Trends and applications in information systems and technologies: Volume 1 - Early prediction of student's performance in higher education: A case study. *Advances in Intelligent Systems Computing,* Volume 1, pp. 166-175.

Martorell, P., Stange, K. & McFarlin, I., 2016. Investing in schools: capital spending, facility conditions, and student achievement. *Journal of Public Economics,* Issue 140, pp. 13-29.

Mason, M., 2008. Complexity theory and the philosophy of education. *Philosophy of Education Society of Australasia,* 40(1), pp. 4-18.

Mason, M., 2016. Is thorough implementation of policy change in education actually possible? What complexity theory tells us about initiating and sustaining change. *European Journal of Education,* 51(4), pp. 437-440.

Maxwell, L. E., 2016. School building condition, social climate, student attendance and academic achievement: A mediation model. *Journal of Environmental Psychology,* Volume 46, pp. 206-216.

McKnight, L. & Morgan, A., 2020. A broken paradigm? What education needs to learn from evidence-based medicine. *Journal of Education Policy,* 35(5), pp. 648-664.

McKnight, L. & Whitburn, B., 2020. Seven reasons to question the hegemony of Visible Learning. *Discourse: Studies in the Cultural Politics of Education,* 41(1), pp. 32-44.

Mendell, M. J. & Heath, G. A., 2005. Do indoor pollutants and thermal conditions in schools influence student performance?: A critical review of the literature. *Indoor Air,* 15(1), pp. 27-52.

Minarechová, M., 2012. Negative impacts of high-stakes testing. *Journal of Pedagogy,* 1(3), pp. 82-100.

Mitchell, M., 2009. *Complexity: A guided tour.* 1st edition ed. Oxford, UK: Oxford University Press.

Mitchell, M., Bradshaw, C. & Leaf, P., 2010. Student and teacher perceptions of school climate: A multilevel exploration of patterns of discrepancy. *Journal of School Health,* 80(6), pp. 271-279.

Mockler, N., 2022. Teacher professional learning under audit: Reconfiguring practice in an age of standards. *Professional Development in Education,* 48(1), pp. 166-180.

Morgan, D. L., 2014. Pragmatism as a paradigm for social research. *Qualitative Inquiry,* 20(8), pp. 1045-1053.

Morgan, H., 2014. The education system in Finland: A success story other countries can emulate. *Childhood Education,* 90(6), pp. 453-457.

Morin, E., 2005. *Restricted complexity, general complexity.* Cerisy-La-Salle, France, “Intelligence de la complexit´e : ´epist´emologie et pragmatique”.

Morrison, K., 2010. Complexity theory, school leadership and management: Questions for theory and practice. *Educational Management Administration and Leadership,* 38(3), pp. 374-393.

Morrison, K., 2021. Taming randomized controlled trials in education: Exploring key claims, issues and debates. 1st edition ed. New York, NY: Routledge.

Morsy, L., Khavenson, T. & Carnoy, M., 2018. How international tests fail to inform policy: The unsolved mystery of Australia's steady decline in PISA scores. *International Journal of Educational Development,* Volume 60, pp. 60-79.

Mughnyanti, M., Efendi, S. & Zarlis, M., 2020. Analysis of determining centroid clustering x-means algorithm with Davies-Bouldin index evaluation. *IOP Conference series: Materials science and engineering,* pp. 1-7.

Nason, R., 2017. *It's not complicated: The art and science of complexity in business.* 1st edition ed. Toronto: University of Toronto Press.

Natekin, A. & Knoll, A., 2013. Gradient boosting machines: A tutorial. *Frontiers in Neurobiotics,* Volume 7, pp. 1-21.

National School Climate Center, 2021. *NSCC: National School Climate Center.* [Online]
Available at: schoolclimate.org
[Accessed 15 October 2024].

National School Improvement Partnerships, 2023. *National School Improvement Partnerships.* [Online]
Available at: https://www.nsipartnerships.com.au/
[Accessed 15 March 2023].

Newman, J., Cherney, A. & Head, B., 2017. Policy capacity and Evidence Based Policy in the public service. *Public Management Review,* 19(2), pp. 157-174.

Nhat-Duc, H. & Van-Duc, T., 2023. Comparison of histogram-based gradient boosting classification machine, random forest, and deep convolutional neural network for pavement raveling severity classification. *Automation in Construction,* Volume 148, pp. 1-16.

NSW Department of Education, 2017. *School Excellence Framework,* Sydney: NSW Department of Education.

NSW Department of Education, 2019. *NSW Department of Education Strategic Plan.* [Online]
Available at: https://education.nsw.gov.au/content/dam/main-education/about-us/strategies-and-reports/media/documents/NSW-Department-of-Education-Strategic-Plan-2018-2022.pdf
[Accessed 11 10 2020].

NSW Department of Education, 2020. *The School Success Model: Building on the Local Schools, Local Decisions reforms.* [Online]
Available at: file:///Users/darryl/Downloads/The-School-Success-Model.pdf
[Accessed 29 November 2021 2021].

NSW Department of Education, 2020. What works best: 2020 update: Overview for parents and carers of primary and secondary school students, Sydney: NSW Department of Education.

NSW Department of Education, 2022. Approval letter to University of Technology, Ethics Secretariat. Sydney: NSW Government.

NSW Department of Education, 2023. *About the School Excellence Framework.* [Online]
Available at: https://education.nsw.gov.au/about-us/strategies-and-reports/school-excellence-and-accountability/sef-evidence-guide/resources/about-sef
[Accessed 1 June 2023].

NSW Department of Education, 2023. *School Success Model explained.* [Online]
Available at: https://education.nsw.gov.au/public-schools/school-success-model/school-success-model-explained
[Accessed 15 June 2023].

NSW Department of Education, 2023. *Tell Them From Me student survey.* [Online]
Available at: https://education.nsw.gov.au/schooling/parents-and-carers/tell-them-from-me/about-tell-them-from-me-/student-survey#About0
[Accessed 17 January 2023].

NSW Department of Education, 2024. *Rebuilding Public Education.* [Online]
Available at: https://education.nsw.gov.au/news/latest-news/nsw-budget--rebuilding-public-education
[Accessed 15 October 2024].

NSW Department of Education, July 2017. *School Excellence Framework Evidence Guide.* [Online]
Available at: https://education.nsw.gov.au/teaching-and-learning/school-excellence-and-accountability/sef-evidence-guide/resources/about-sef
[Accessed 10 November 2021].

NSW Government, 1998. *Children and Young Persons (Care and Protection) Act 1998 (NSW).* [Online]
Available at: https://legislation.nsw.gov.au/view/html/inforce/current/act-1998-157
[Accessed 10 January 2025].

NSW Government, 1998. *Privacy and Personal Information Protection Act 1998 (NSW)*. [Online]
Available at: https://legislation.nsw.gov.au/view/html/inforce/current/act-1998-133
[Accessed 10 January 2025].

NSW Government, 2023. *2022-23 Half Yearly Review.* [Online]
Available at: https://www.budget.nsw.gov.au/my-budget
[Accessed 25 June 2023].

OECD, 2018. Education policy outlook 2018: Putting student learning at the centre, Paris: OECD Publishing.

OECD, 2020. Education at a glance 2020: OECD indicators, Paris: OECD Publishing.

Ofem, U. J. et al., 2024. Predicting affective and cognitive learning outcomes: A quantitative analysis using climate change vectors. *Eurasian Journal of Science and Environmental Education,* 4(1), pp. 1-12.

O'Neill, D. J. & Oates, A. D., 2001. The impact of school facilities on student achievement, behavior, attendance, and teacher turnover rate in Central Texas middle schools. *Educational Facility Planner,* 36(3), pp. 14-22.

Ojha, R., 2023. A case of fresh fruits and vegetables supply chain performance improvement: A system dynamics modelling and analysis. *International Journal of Logistics Economics and Globalisation,* 10(1), pp. 23-41.

Oliver, K., 2022. How policy appetites shape and are shaped by evidence production and use. In: P. Fafard, A. Cassola & E. de Leeuw, eds. *Integrating science and politics for public health.* Cham: Palmgrave Macmillan, pp. 77-101.

O'Neill, J., 2012. Material fallacies of education research evidence and public policy advice. *New Zealand Journal of Educational Studies,* 47(1), pp. 153-159.

Paivio, A., 1990. *Mental representations: A dual coding approach.* Oxford: Oxford University Press.

Parker, P. D. et al., 2018. Inequity and excellence in academic performance: Evidence from 27 countries. *American Educational Research Journal,* Volume 55, pp. 836-858.

Parkhurst, J., 2017. The politics of evidence: From Evidence-Based Policy to the good governance of evidence.. 1st edition ed. Abingdon, UK: Routledge.

Pelleg, D. & Moore, A., 2000. X-means: Extending k-means with efficient estimation of the number of clusters. Pittsburgh, Carnegie Mellon University.

Pezoulas, V. C. et al., 2024. FHBF: Federated hybrid boosted forests with dropout rates for supervised learning tasks across highly imbalanced clinical datasets. *Patterns,* 5(1), pp. 1-14.

Polesel, J., Rice, S. & Dulfer, N., 2014. The impact of high-stakes testing on curriculum and pedagogy: A teacher perspective from Australia. *Journal of Education Policy,* 29(5), pp. 640-657.

Portfolio Committee No. 3 – Education, 2020. *Measurement and outcome-based funding in New South Wales schools,* Sydney: New South Wales Parliament.

Prigogine, I. & Stengers, I., 1984. *Title Order out of chaos : Man's new dialogue with nature.* 1st edition ed. London: Heinemann.

Putansu, S., 2020. Politics and policy knowledge in federal education : Confronting the evidence-based proverb. 1st edition ed. Cham, Switzerland: Springer International.

Raines, D. A., 2013. Research paradigms and methods.. *Neonatal Network,* 32(6), pp. 425-428.

Rapidminer, 2023. *Rapidminer.* [Online]
Available at: https://rapidminer.com/
[Accessed 1 September 2023].

Reid, A., 2020. Changing Australian education: How policy is taking us backwards and what can be done about it. New York: Allen and Unwin.

Reimers, F., 2009. "Global competency" is imperative for global success. *The Chronicle of Higher Education,* 1(55), pp. 1-4.

Renshaw, T. & Chenier, J., 2016. Further validation of the Student Subjective Wellbeing Questionnaire: Comparing first-order and second-order factor effects on actual school outcomes. *Journal of Psychoeducational Assessment,* 36(4), pp. 392-397.

Reupert, A., 2019. Mental health and academic learning in schools : Approaches for facilitating the wellbeing of children and young people. Oxford: Taylor and Francis Group.

Reynolds, K. J. et al., 2017. How does school climate impact academic achievement?: An examination of social identity processes. *School Psychology International,* 38(1), pp. 78-97.

Roberts, L. W., 2009. Measuring school facility conditions: An illustration of the importance of purpose. *Journal of Educational Administration,* 47(3), pp. 368-380.

Rogers, K., 2020. The effects of classroom seating layouts on participation and assessment performance in a fourth grade classroom. *Journal of Learning Spaces,* 9(1), pp. 31-41.

Romero, C. & Ventura, S., 2010. Educational data mining: A review of the state of the art. *IEEE Transactions on Human-Machine Systems,* 40(6), pp. 601-618.

Roots, R., 2004. When laws backfire: Unintended consequences of public policy. *The American Behavioral Scientist,* 47(11), pp. 1376-1394.

Ross, C. & Wu, C., 1995. The links between education and health. *American Sociological Review,* Volume 60, pp. 719-745.

Rudasill, K., Snyder, K., Levinson, H. & Adelson, J., 2018. Systems view of school climate: A theoretical framework for research. *Educational Psychology Review,* 30(1), pp. 35-60.

Ruggiero, J., Miner, J. & Blanchard, L., 2002. Measuring equity of educational outcomes in the presence of inefficiency. *European Journal of Operational Research,* 142(3), pp. 642-652.

Sahlberg, P., 2007. Secondary education in OECD countries: Common challenges, differing solutions, Torino: European Training Foundation.

Sahlberg, P., 2021. Finnish Lessons 3.0: What can the world learn from educational change in Finland?. New York: Teachers College Press.

Sanderson, I., 2009. Intelligent policy making for a complex world: Pragmatism, evidence and learning. *Political Studies, 57*(4), pp. 699-719.

Saunders, M., Lewis, P. & Thornhill, A., 2016. *Research methods for business students.* 7th edition ed. Essex, United Kingdom: Pearson.

Sayer, P., 2020. A new epistemology of evidence-based policy. *Policy and Politics,* 48(2), pp. 241-258.

Schmid, C., Trendtel, M., Bruneforth, M. & Hartig, J., 2020. Effectiveness of a governmental action to improve Austrian primary schools: Results of multilevel analyses based on repeated cycles of educational standards assessments. *School Effectiveness and School Improvement,* 31(2), pp. 149-171.

Sellar, S. & Lingard, B., 2014. Equity in Australian schooling: The absent presence of socioeconomic context. In: M. Somerville, ed. *Contemporary Issues of Equity in Education.* Cambridge: Cambridge Scholars, pp. 1-20.

Sellström, E. & Bremberg, S., 2006. Is there a "school effect" on pupil outcomes?: A review of multilevel studies. *Journal of Epidemiology and Community Health,* Volume 60, pp. 149-155.

Sheeran, T. J. & Sheeran, M. F., 1996. Schools, schooling, and teachers: A curriculum for the future. *NASSP Bulletin,* Volume 80, pp. 47-56.

Shield, B., Greenland, E. & Dockrell, J., 2010. Noise in open plan classrooms in primary schools: A review. *Noise and Health,* 12(49), pp. 225-234.

Shook, J. R., 2023. *Pragmatism.* 1st edition ed. Cambridge, Massachusetts: The MIT Pres.

Shrestha, Y. R., He, V. F., Puranam, P. & von Krogh, G., 2021. Algorithm supported induction for building theory: How can we use prediction models to theorize?. *Organization Science,* 32(3), pp. 856-880.

Shukla, K., Konold, T. & Cornell, D., 2016. Profiles of student perceptions of school climate: Relations with risk behaviors and academic outcomes. *American Journal of Community Psychology,* 57(3), pp. 291-307.

Sianesi, B. & Reenen, J. V., 2003. The returns to education: Macroeconomics. *Journal of Economic Surveys,* 17(2), pp. 157-200.

Simpson Reeves, S., Exley, B. & Dillon-Wallace, J., 2018. Secondary school English teachers caught in the NAPLAN fray: Effects of the disparate responses. *English in Australia,* 53(1), pp. 24-32.

Simpson, A., 2017. The misdirection of public policy: Comparing and combining standardised effect sizes. *Journal of Education Policy,* 32(4), pp. 450-466.

Simpson, A., 2018. Princesses are bigger than elephants: Effect size as a category error in evidence-based education. *Journal of the British Education Research Association,* Volume 44, pp. 897-9132.

Slavin, R., 2008. Perspectives on evidence-based research in education. *Educational Researcher,* 37(1), pp. 5-14.

Sleegers, P. J. C. et al., 2013. Lighting affects students' concentration positively: Findings from three Dutch studies. *Lighting Research and Technology,* 19(3), pp. 359-368.

Snook, I. et al., 2009. Invisible learnings?: A commentary on John Hattie's book Visible Learning: A synthesis of over 800 meta-analyses relating to achievement, Palmerston North: Massey University.

Snyder, S., 2013. The simple, the complicated, and the complex: Educational reform through the lens of complexity theory, s.l.: OECD.

Spataru, C., 2015. Whole energy system dynamics: Theory, modelling, and policy. New York, NY: Routledge.

Steen, M. v. D., Twist, M. v., Fenger, M. & Cointre, S. L., 2013. Complex causality in improving underperforming schools: A complex adaptive systems approach. *Policy Politics,* Volume 41, pp. 551-567.

Sterman, J. D., 2000. Business dynamics: Systems thinking and modeling for a complex world. San Francisco: Irwin McGraw-Hill.

Swinton, O. H., 2010. The effect of effort grading on learning. *Economics of Education Review,* 29(6), pp. 1176-1182.

Tan, P. N., Steinbach, M. & Kumar, V., 2019. *Introduction to Data Mining.* 2nd edition ed. London: Pearson.

Tashakkori, A. & Teddlie, C., 2010. *The Sage handbook of mixed methods in social and behavioural research.* 2nd edition ed. Thousand Oaks, California: Sage.

Taylor, J. & Alanzi, S., 2023. Cohen's and Hedges' g. *The Journal of Nursing Education,* 62(5), pp. 316-317.

Terano, T., 2005. Agent-based simulation : From modeling methodologies to real-world applications : Post-proceedings of the third international workshop on agent-based approaches in economic and social complex systems 2004. New York, Springer.

Terhart, E., 2011. Has John Hattie really found the holy grail of research on teaching?: An extended review of Visible Learning. *Journal of Curriculum Studies,* 43(3), pp. 425-438.

Thapa, A., Cohen, J., Guffey, S. & Higgins-D'Alessandro, A., 2013. A review of school climate research. *Review of Educational Research,* 83(3), pp. 357-385.

The Learning Bar, 2024. *Tell Them From Me.* [Online]
Available at: https://nsw.tellthemfromme.com/main.htm
[Accessed 24 April 2024].

Tin, T., Hock, L. & Ikumapayi, O., 2024. Educational big data mining: Comparison of multiple machine learning algorithms in predictive modelling of student academic performance. *Educational Big Data Mining,* 15(6), pp. 633-645.

Totfum, J. et al., 2015. Association between classroom ventilation mode and learning outcome in Danish schools. *Building and Environment,* Volume 92, pp. 494-503.

Touzani, S., Granderson, J. & Fernandes, S., 2018. Gradient boosting machines for modeling the the energy consumption of commercial buildings. *Energy and Buildings,* Volume 158, pp. 1522-1543.

Toyinbo, O., Jengo, E., Peralta, X. V. & Habler, B., 2024. Indoor environmental quality and student comfort in East Africa schools: Insights from the improving learning through classroom experience. *SSRN,* pp. 1-24.

Trifonova, O. P., Lokhov, P. G. & Archakov, A. I., 2013. Metabolic profiling of human blood. *Biochemistry,* 7(3), pp. 179-186.

Uline, C., Devere Wolsey, T., Tschannen-Moran, M. & Lin, C., 2010. Improving the physical and social environment of school: A question of equity. *Journal of School Leadership,* Volume 20, pp. 597-632.

Uline, C. & Tschannen-Moran, M., 2008. The walls speak: The interplay of quality facilities, school climate, and student achievement. *The Journal of Educucation Administration,* Volume 46, pp. 55-73.

United Nations. Department of Economic and Social Affairs, 2024. *SDG 4: Education: Progress and Info.* [Online]
Available at: https://sdgs.un.org/goals/goal4#progress_and_info
[Accessed 2 November 2024].

United Nations, 2024. *Sustainable Development Goals.* [Online]
Available at: https://www.un.org/sustainabledevelopment/
[Accessed 1 September 2024].

Urick, A. & Bowers, A. J., 2014. The impact of principal perception on student academic climate and achievement in high school: How does it measure up?. *Journal of School Leadership,* Volume 24, pp. 386-414.

Vakil, M. K., Khan, A. & Rahman, A., 2022. *Analysing the academic performance of students under varying climatic conditions.* 1st edition ed. Singapore: Springer Nature.

Van Claster, B. et al., 2019. Calibration: The Achilles heel of predictive analytics. *BMC Medicine,* Volume 17, pp. 2-7.

Van Houtte, M. & Van Maele, D., 2011. The black box revelation: In search of conceptual clarity regarding climate and culture in school effectiveness research. *Oxford Review of Education,* 37(4), pp. 505-524.

van Vondel, S., Steenbeck, H., van Dijk, M. & van Geert, P., 2016. "Looking at" educational interventions: Surplus value of a complex dynamic systems approach to study the effectivenessof a science and technology educational intervention. In: M. S. D. Koopmans, ed. *Complex and dynamical systems in education.* Cham: Springer, pp. 203-232.

Velasco-Benítez, C. A. & García-Perdomo, H. A., 2019. Some flaws that could change the effect sizes and limit the extrapolation of these results. *Jornal de Pediatria,* 95(3), pp. 374-375.

Venter, L., 2022. A systems perspective on early childhood development education in South Africa. *International Journal of Child Care and Education Policy,* 16(1), pp. 1-25.

Villar, A. & Andrade, C. R. V., 2023. Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study. *Discoveries in Artificial Intelligence,* 4(2), pp. 1-25.

Vulic, J., Jacobson, M. J. & Levin, J. A., 2024. Exploring education as a complex system: Computational educational research with multi-level agent-based modeling. *Education Sciences,* 14(5), p. 551.

Waltenberg, F. D. & Vandenberghe, V., 2007. What does it take to achieve equality of opportunity in education?: An empirical investigation based on Brazilian data. *Economics of Education Review,* 26(6), pp. 709-723.

Wang, M. T. & Degol, J. L., 2016. School climate: a Review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review,* Volume 28, pp. 315-352.

Wargocki, P. & Wyon, D. P., 2013. The effects of moderately raised classroom temperatures and classroom ventilation rate on the performance of schoolwork by children. *HVAC Research,* 19(3), pp. 359-368.

Warren, K., 2015. *Strategy dynamics essentials.* 2nd edition ed. Princes Risborough: Strategy Dynamics.

Waslander, S., Hooge, E. & Drewes, T., 2016. Steering dynamics in the Dutch education system. *European Journal of Education,* 51(4), pp. 478-494.

Waslander, S., Hooge, E. H., Theisens, H. C. & Pater, C. J., 2020. Steering in complex education systems: Why similar aims can have dramatically different results. *Journal of Education Policy,* 35(2), pp. 168-187.

Williams, J. & Enge, L. C., 2012. How do other countries evaluate teachers?. *Phi Delta Kappan,* 94(4), pp. 53-57.

Wiseman, A. & Davidson, P., 2018. *Cross-nationally comparative, evidence-based educational policymaking and reform.* 1st edition ed. Bongley, UK: Emerald Publishing.

Wong, A., Lai, C., Shum, A. & Yip, P., 2022. From the hidden to the obvious: Classification of primary and secondary student suicides using cluster analysis. *BMC Public Health,* Volume 22.

World Bank Group, 2025. *Education Statistics (EdStats).* [Online]
Available at: https://datatopics.worldbank.org/education/
[Accessed 15 Feb 2025].

World Population Review, 2024. *PISA scores by countries 2024.* [Online]
Available at: https://worldpopulationreview.com/country-rankings/pisa-scores-by-country
[Accessed 27 December 2024].

Wright, D. & Meadows, D. H., 2009. *Thinking in systems: A primer.* 1st edition ed. London, U.K.: Routledge.

Wrigley, T., 2019. The problem of reductionism in educational theory: Complexity, causality, values. *Power and Education,* 11(2), pp. 145-162.

Wu, J. & Coggeshall, S., 2012. *Foundations of predictive analytics.* 1st edition ed. Boca Raton: CRC Press.

Xu, Z., Huang, G., Weinberger, K. Q. & Zheng, A. X., 2014. *Gradient boosted feature selection..* New York, 20th ACM SIGKDD international conference on knowledge discovery and data mining.

Yang, Y., 2024. Complexity theory for complexity reduction?: Revisiting the ontological and epistemological basis of complexity science with Critical Realism. *Journal for the Theory of Social Behaviour,* 54(4), pp. 368-383.

Yang, Y., Wu, Q. & Ma, X., 2020. Enhancing small sample learning with gradient boosted trees. *Pattern Recognition Letters,* Volume 138, pp. 10-20.

Yorke, M., 2009. 'Student experience' surveys: Some methodological considerations and an empirical investigation. *Assessment and Evaluation in Higher Education,* 34(6), pp. 721-739.

Yu, A. & Ho, C., 2022. Ensemble methods: The wisdom of the crowd. In: *Data mining and exploration: From traditional statistics to modern data science.* s.l.:Taylor and Francis Group, pp. 153-178.

Yurkofsky, M. M. et al., 2020. Research on continuous improvement: Exploring the complexities of managing educational change. *Review of Research in Education,* March, 44(1), pp. 403-433.

Zabek, F. et al., 2022. Can a school climate survey accurately and equitably measure school quality?: Examining the multilevel structure and invariance of the Georgia School Climate Scale. *Journal of School Psychology,* Volume 95, pp. 1-24.

Zadza, J., 2020. Globalisation, ideology and education reforms : Emerging paradigms. 1st edition ed. Doerdrecht, Netherlands: Springer.

Zaman, B. et al., 2023. Modeling education impact: A machine learning-based approach for improving the quality of school education. *Journal of Computers in Education,* 11(4), pp. 1181-1214.

Zawacki-Richter, O., Marín, V. I., Bond, M. & Gouverneur, F., 2019. Systematic review of research on artificial intelligence applications in higher education: Where are the educators?. *International Journal of Educational Technology in Higher Education,* 16(39), pp. 1-27.

Zengaro, S. & Warley, R., 2022. The fallacy of higher education as the great social equalizer: Racial identity, implicit bias, and achievement. In: J. Keengwer, ed. *Handbook of research on social justice and equity in education.* Hershey, PA: IGI Global, pp. 46-71.

Zullig, J., Koopman, T., Patton, J. & Ubbes, V., 2010. School climate: Historical review, instrument development, and school assessment. *Journal of Psychoeducational Assessment,* 28(2), pp. 139-151.

Appendices

Appendix 1. Survey instrument

WHAT'S HAPPENING IN THIS SCHOOL (WHITS-S)

SECONDARY SCHOOL

NEW SOUTH WALES: PILOT

School:	

Background Information

i.	Gender	Male	Female	Indeterminate/intersex/ unspecified		Other	
ii.	If other, please specify	Prefer not to say	Transgender female	Transgender male	Non-binary	Not sure	
iii.	What school level are you in?	Middle school (years 7, 8 & 9)			Senior school (years 10, 11 & 12)		
iv.	What grade are you in?	7	8	9	10	11	12
v.	Is English the first language that you learned to speak?	Yes			No		
vi.	If yes, do you speak a language other than English at home?	Yes			No		
vii.	Where were you born?	Australia			Overseas		
viii.	Are you Aboriginal or Torres Strait Islander?	Yes		No		Prefer not to say	
ix.	Do you have a disability (e.g., Autism, ADHD, mobility impairment, intellectual disability, dyslexia, or other)?	Yes		No		Prefer not to say	
	If yes, please choose the most relevant one:	☐ Prefer not to say ☐ My disability is physical ☐ My disability affects my learning ☐ My disability affects how I feel in different environments ☐ My disability affects how I relate to the people around me ☐ My disability affects several or all of the above					

Safety

Knowing where to go for help	*ACTUAL*					*PREFERRED*				
	Almost Never	*Not Often*	*Sometimes*	*Often*	*Almost Always*	*Almost Never*	*Not Often*	*Sometimes*	*Often*	*Almost Always*
1. If I told a teacher that I was being bullied, they would help.	1	2	3	4	5	1	2	3	4	5
2. If a student were bullying me, I would report it to a teacher.	1	2	3	4	5	1	2	3	4	5
3. I can report incidents without others finding out.	1	2	3	4	5	1	2	3	4	5
4. If I felt unsafe, I would tell an adult at the school.	1	2	3	4	5	1	2	3	4	5
5. Someone in a position of responsibility is available to support me if I need it.	1	2	3	4	5	1	2	3	4	5

Clear rules	*ACTUAL*					*PREFERRED*				
	Almost Never	*Not Often*	*Sometimes*	*Often*	*Almost Always*	*Almost Never*	*Not Often*	*Sometimes*	*Often*	*Almost Always*
1. The rules at this school are reasonable.	1	2	3	4	5	1	2	3	4	5
2. The rules make it clear that certain behaviours are not okay.	1	2	3	4	5	1	2	3	4	5
3. The consequences for breaking the school rules are clear.	1	2	3	4	5	1	2	3	4	5
4. The school rules are applied equally to all students.	1	2	3	4	5	1	2	3	4	5
5. If I break the rules, the school staff will help me learn from my mistake.	1	2	3	4	5	1	2	3	4	5

Feeling safe	ACTUAL					PREFERRED				
	Almost Never	Not Often	Sometimes	Often	Almost Always	Almost Never	Not Often	Sometimes	Often	Almost Always
1. I feel safe before and after school while on school grounds.	1	2	3	4	5	1	2	3	4	5
2. I feel safe during break times.	1	2	3	4	5	1	2	3	4	5
3. I feel safe using the toilet facilities at school.	1	2	3	4	5	1	2	3	4	5
4. I feel safe in the learning spaces in the school (e.g., classrooms).	1	2	3	4	5	1	2	3	4	5
5. I feel safe in in the outdoor areas of the school.	1	2	3	4	5	1	2	3	4	5

Community

Teachers	ACTUAL					PREFERRED				
	Almost Never	Not Often	Sometimes	Often	Almost Always	Almost Never	Not Often	Sometimes	Often	Almost Always
1. Teachers try to understand my problems.	1	2	3	4	5	1	2	3	4	5
2. Teachers take an interest in my background.	1	2	3	4	5	1	2	3	4	5
3. The teachers support me when I have problems.	1	2	3	4	5	1	2	3	4	5
4. The teachers really care about me.	1	2	3	4	5	1	2	3	4	5
5. Teachers treat me with respect.	1	2	3	4	5	1	2	3	4	5

Friendships	ACTUAL					PREFERRED				
	Almost Never	Not Often	Sometimes	Often	Almost Always	Almost Never	Not Often	Sometimes	Often	Almost Always
1. I belong to a group of friends.	1	2	3	4	5	1	2	3	4	5
2. I make friends with students from different backgrounds.	1	2	3	4	5	1	2	3	4	5
3. I socialise with other students.	1	2	3	4	5	1	2	3	4	5
4. I have friends who care about me.	1	2	3	4	5	1	2	3	4	5
5. If I have a problem, there are students who are willing to help me.	1	2	3	4	5	1	2	3	4	5

Belonging	*ACTUAL*					*PREFERRED*				
	Almost Never	*Not Often*	*Sometimes*	*Often*	*Almost Always*	*Almost Never*	*Not Often*	*Sometimes*	*Often*	*Almost Always*
1. I feel accepted at school.	1	2	3	4	5	1	2	3	4	5
2. I feel included at school.	1	2	3	4	5	1	2	3	4	5
3. I feel part of a community when I am at school.	1	2	3	4	5	1	2	3	4	5
4. I feel respected when I am at school.	1	2	3	4	5	1	2	3	4	5
5. I feel valued when I am at school.	1	2	3	4	5	1	2	3	4	5

Understanding difference	*ACTUAL*					*PREFERRED*				
	Almost Never	*Not Often*	*Sometimes*	*Often*	*Almost Always*	*Almost Never*	*Not Often*	*Sometimes*	*Often*	*Almost Always*
1. Days that are important to my culture are recognised.	1	2	3	4	5	1	2	3	4	5
2. I am encouraged to understand the culture of others.	1	2	3	4	5	1	2	3	4	5
3. My background is known by students and teachers.	1	2	3	4	5	1	2	3	4	5
4. My differences are respected.	1	2	3	4	5	1	2	3	4	5
5. Cultural traditions that are relevant to me are recognised as important.	1	2	3	4	5	1	2	3	4	5

Academic

Making learning interesting	*ACTUAL*					*PREFERRED*				
In general, at this school, …	*Almost Never*	*Not Often*	*Sometimes*	*Often*	*Almost Always*	*Almost Never*	*Not Often*	*Sometimes*	*Often*	*Almost Always*
1. The learning activities are relevant to me.	1	2	3	4	5	1	2	3	4	5
2. The learning activities are of interest to me.	1	2	3	4	5	1	2	3	4	5
3. The learning activities are of practical value.	1	2	3	4	5	1	2	3	4	5
4. I enjoy the learning activities.	1	2	3	4	5	1	2	3	4	5
5. The learning activities motivate me to want to learn.	1	2	3	4	5	1	2	3	4	5

My learning	*ACTUAL*					*PREFERRED*				
	Almost Never	*Not Often*	*Sometimes*	*Often*	*Almost Always*	*Almost Never*	*Not Often*	*Sometimes*	*Often*	*Almost Always*
1. My teachers give me individual attention when I need it.	1	2	3	4	5	1	2	3	4	5
2. My teachers provide work that is important to my future.	1	2	3	4	5	1	2	3	4	5
3. My teachers notice if I have trouble learning something.	1	2	3	4	5	1	2	3	4	5
4. My teachers give me homework that helps me to learn.	1	2	3	4	5	1	2	3	4	5
5. My teachers give me extra help with my schoolwork if I need it.	1	2	3	4	5	1	2	3	4	5

Teacher feedback	*ACTUAL*					*PREFERRED*				
	Almost Never	*Not Often*	*Sometimes*	*Often*	*Almost Always*	*Almost Never*	*Not Often*	*Sometimes*	*Often*	*Almost Always*
1. Feedback from my teachers about my learning helps me understand what I need to do to improve.	1	2	3	4	5	1	2	3	4	5
2. Feedback from assessment tasks helps me to improve my learning.	1	2	3	4	5	1	2	3	4	5
3. Feedback for assessment tasks is at the right level of detail for me.	1	2	3	4	5	1	2	3	4	5
4. Feedback for assessment tasks is returned promptly.	1	2	3	4	5	1	2	3	4	5
5. Feedback about where I went wrong helps me to learn.	1	2	3	4	5	1	2	3	4	5

Institutional Environment

Buildings and grounds - when thinking about the buildings and grounds, how strongly do you agree with these statements?

	Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*
1. The outside of the school buildings are well maintained (upkeep and painting).	1	2	3	4	5
2. The classroom furniture and flooring are well maintained.	1	2	3	4	5
3. The grounds of the school (gardens and other outdoor spaces) are well maintained.	1	2	3	4	5
4. The inside of the school (e.g., classrooms, walkways, lockers) are well maintained.	1	2	3	4	5
5. The school is kept clean and tidy.	1	2	3	4	5

The feel of my classrooms - when thinking about the feel of your classrooms, how strongly do you agree with these statements?					
Think about the different spaces where you learn …	*Strongly Disagree*	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*
1. The temperature is about right (neither too hot nor too cold).	1	2	3	4	5
2. There is a pleasant smell or no smell.	1	2	3	4	5
3. The quality of light is about right (neither too bright nor too dark).	1	2	3	4	5
4. During study or lesson time, I am not disturbed by noises from outside or inside (e.g., the air conditioner).	1	2	3	4	5
5. I can see what is displayed (e.g., screens, whiteboard) without difficulty.	1	2	3	4	5

Comfort - when thinking about the furniture at school, how strongly do you agree with these statements?					
	Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*
1. I can choose a chair or adjust its height to be the right height for me.	1	2	3	4	5
2. I can choose a table/desk or adjust its height to be the right height for me.	1	2	3	4	5
3. I can adjust the height of my computer screen to be the right height for me.	1	2	3	4	5
4. Desks and/or chairs are fitted with castor wheels.	1	2	3	4	5
5. The chairs and other seating are comfortable to sit on.	1	2	3	4	5

About my classrooms - when thinking about the design of the classrooms, how strongly do you agree with these statements?					
	Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*
1. The design of the school meets my learning needs.	1	2	3	4	5
2. The design of the school provides welcoming learning spaces.	1	2	3	4	5
3. The design of the school makes sure that movement does not get blocked or overcrowded (e.g., along pathways and around locker areas).	1	2	3	4	5
4. The learning spaces (not including classrooms) are adequate for different subjects (e.g., drama, science, music, sports).	1	2	3	4	5
5. In general, I am satisfied with the learning spaces at the school.	1	2	3	4	5

Design of Outdoor Spaces - when thinking about the design of the outside spaces, how strongly do you agree with these statements?					
	Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*
1. There are sufficient outdoor spaces (e.g., playgrounds, grassed areas, ovals).	1	2	3	4	5
2. There is permanent equipment suitable for all ages (such as, climbing frames, obstacles).	1	2	3	4	5
3. There are outdoor spaces I can go to relax or for quiet reflection.	1	2	3	4	5
4. There are outdoor spaces where I can work or socialise with my peers.	1	2	3	4	5
5. There are outdoor spaces where I can play games with my peers.	1	2	3	4	5

Adequacy of the Outdoor Spaces - thinking about the how appropriate the outdoor spaces are for you, how strongly do you agree with these statements?					
	Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*
1. There is enough space for the number of students that use the outdoor spaces.	1	2	3	4	5
2. There is enough outside equipment (such as basketball hoops) for all who want to use it.	1	2	3	4	5
3. There are enough hard ball courts (e.g., tennis and basketball) for all who want to use them.	1	2	3	4	5
4. There are enough sports fields (grassed areas such as ovals).	1	2	3	4	5
5. There is enough shelter for protection from the wind, rain and sun.	1	2	3	4	5

Institutional Environment – Additional Questions

Use of Space - Are the following spaces are available in your school?				
1. A traditional classroom with no direct access to other spaces (such as areas for group work, individual work etc.).	Yes		No	
If yes, during lesson time, how often have you used this space over the last week?	Never	Once	2 to 4 times	Every day
2. A traditional classroom with direct access to other spaces (e.g., for collaborative group work, project work or individual work).	Yes		No	
If yes, during lesson time, how often have you used this space over the last week?	Never	Once	2 to 4 times	Every day
3. A teaching area where two or more teachers can work together and share connected learning spaces (e.g., for collaborative group work, project work or individual work).	Yes		No	
If yes, during lesson time, how often have you used this space over the last week?	Never	Once	2 to 4 times	Every day
4. A library.	Yes		No	
If yes, during lesson time, how often have you used this space over the last week?	Never	Once	2 to 4 times	Every day

5. A hall/auditorium.	Yes		No		
If yes, during lesson time, how often have you used this space over the last week?	Never	Once	2 to 4 times	Every day	
6. A canteen.	Yes		No		
If yes, during lesson time, how often have you used this space over the last week?	Never	Once	2 to 4 times	Every day	
7. A science laboratory.	Yes		No		
If yes, during lesson time, how often have you used this space over the last week?	Never	Once	2 to 4 times	Every day	
8. A workshop/studio space for art, music or design.	Yes		No		
If yes, during lesson time, how often have you used this space over the last week?	Never	Once	2 to 4 times	Every day	
9. A kitchen/food technology space.	Yes		No		
If yes, during lesson time, how often have you used this space over the last week?	Never	Once	2 to 4 times	Every day	
10. A workshop space for technology (e.g., wood, metal, plastics, robotics).	Yes		No		
If yes, during lesson time, how often have you used this space over the last week?	Never	Once	2 to 4 times	Every day	
11. A gym.	Yes		No		
If yes, during lesson time, how often have you used this space over the last week?	Never	Once	2 to 4 times	Every day	
12. Are there other spaces available in your school?	Yes		No		
If yes, please briefly describe them here:					

Use of Space - Think of the external (outside) spaces in your school that you can get to straight from a classroom? Do you have any of the following:					
1. An external (outside) classroom or space—usually with seating.	Yes		No		
If yes, during lesson time, approximately how often do you use these external (outside) spaces in your school over the current school year?	Never	Hardly ever	At least once a month	At least once a week	Every day
2. A sports field.	Yes		No		
If yes, during lesson time, approximately how often do you use these external (outside) spaces in your school over the current school year?	Never	Hardly ever	At least once a month	At least once a week	Every day
3. A schoolyard.	Yes		No		
If yes, during lesson time, approximately how often do you use these external (outside) spaces in your school over the current school year?	Never	Hardly ever	At least once a month	At least once a week	Every day
4. Are there other external (outside) spaces, directly accessible from a classroom, in your school?	Yes		No		
If yes, please briefly describe them here:					

Buildings	
1. Thinking about the different spaces in which you learn, state which learning spaces are inadequate and why?	

Protective and Risk Factors

Self-belief					
	Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*
1. At school, even if the work is hard, I can learn it.	1	2	3	4	5
2. I feel that I will achieve a good result at school.	1	2	3	4	5
3. At school, I can do difficult work if I try.	1	2	3	4	5
4. I understand what is taught at school.	1	2	3	4	5

You will be asked to indicate the frequency of which you may have experienced conflict or social harm at your school. This content is disturbing. NSI encourages you to prepare emotionally before responding to these items. If you believe that responding to these items will be traumatizing, please select that you do not wish to respond.

- ☐ I wish to respond
- ☐ I do not wish to respond

If you selected that you do not wish to respond to statements related to your experience of conflict or social harm, please skip to page 11.

Conflict					
At this school, in the last few weeks ...	*Never*	*Hardly ever*	*At least once a month*	*At least once a week*	*Every day*
1. I have been verbally harassed.	1	2	3	4	5
2. I have been physically harassed.	1	2	3	4	5
3. I have been bullied online or on social media.	1	2	3	4	5
4. I have been picked on.	1	2	3	4	5

Vigour

	Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*
1. I feel full of energy when I am at school.	1	2	3	4	5
2. I feel bubbly and full of life when I am at school.	1	2	3	4	5
3. I feel wide awake when I am at school.	1	2	3	4	5
4. I look forward to each new day when I come to school.	1	2	3	4	5

Ability to bounce back

In general, ...	*Strongly Disagree*	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*
1. I am determined to achieve my goals.	1	2	3	4	5
2. I bounce back after difficult times.	1	2	3	4	5
3. I can achieve goals despite barriers.	1	2	3	4	5
4. I come through difficult times with little trouble.	1	2	3	4	5

My learning behaviour - How often is this true?

	Almost Never	*Not Often*	*Sometimes*	*Often*	*Almost Always*
1. I try hard to do well when I am in class.	1	2	3	4	5
2. I work as hard as I can when I am in class.	1	2	3	4	5
3. I pay attention when I am in class.	1	2	3	4	5
4. I listen carefully when I am in class.	1	2	3	4	5

Disruptive behaviour - How often have these things happened this term?					
	Almost Never	*Not Often*	*Sometimes*	*Often*	*Almost Always*
1. Students have used their phones/tablets/devices in class for non-educational purposes.	1	2	3	4	5
2. Students have been disrespectful to teachers.	1	2	3	4	5
3. Students have misbehaved in class.	1	2	3	4	5
4. Students have ignored the teachers' instructions.	1	2	3	4	5

Risky Behaviours					
At school, in the last few weeks...	*Never*	*Once*	*A few times*	*Regularly*	*A lot*
1. I smoked cigarettes/e-cigarettes or have vaped.	1	2	3	4	5
2. I drank alcohol.	1	2	3	4	5
3. I used drugs.	1	2	3	4	5
4. I damaged school property on purpose.	1	2	3	4	5
5. I skipped lessons or chose not to come to school.	1	2	3	4	5
6. I have visited inappropriate websites.	1	2	3	4	5

Other questions

1. After completing school, I intend to ...
   - ☐ Start work.
   - ☐ Take a gap year.
   - ☐ Complete work experience or an internship.
   - ☐ Complete a vocational or trade qualification (VET).
   - ☐ Complete a TAFE course or diploma.
   - ☐ Complete a University degree.
2. Have you been made to feel welcome at this school?
   - ☐ I have always felt welcome.
   - ☐ I have felt unwelcome or uncomfortable.

   Which of the following most applies to you? I have been made to feel unwelcome because of ...
   - ☐ This question does not apply to me.
   - ☐ My gender.
   - ☐ My sex (male/female).
   - ☐ My ethnic background.
   - ☐ My grades.
   - ☐ My appearance.
   - ☐ My religion.
   - ☐ My family's income level.
   - ☐ My disability.
   - ☐ My sexual orientation.
   - ☐ Other.

   If other, please specify:
3. There is at least one adult at this school who cares about what happens to me.
   - ☐ Yes.
   - ☐ No.
   - ☐ Maybe.

4. There is at least one adult at this school that I can trust.	☐ Yes. ☐ No. ☐ Maybe.

Appendix 2. Approval to use survey data for PhD research

NSW GOVERNMENT | Education

15 September 2022

University of Technology, Ethics Secretariat

Dear University of Technology, Ethics Secretariat,

The NSW Department of Education employs Darryl Walker as Director of Research and Innovation in School Infrastructure New South Wales (SINSW). In this capacity, Darryl is leading a research project to understand better how infrastructure can be managed for improved education outcomes.

In recent years, NSW Education has invested significantly in infrastructure. The reasons for this investment are twofold. First, many of the infrastructure assets were reaching the point where they required refurbishment. Second, a growing and shifting population base strained existing capacity.

The need for investment in infrastructure is ongoing, with an increasing emphasis on improving education outcomes. International research shows that, at the macro level, investment in school infrastructure improved education outcomes. Unfortunately, the existing research does not assist decision-makers at the school level; yet this is where investments deliver education outcomes.

I understand that Darryl is also researching his PhD in this area. Specifically, Darryl seeks to use the results of a School Climate Survey combined with Systems Dynamics Modelling to forecast and investigate the effects of infrastructure changes on School Climate.

This research, specifically the collection and use of School Climate Data from students, parents and teachers, is endorsed by the Department. Whilst Darryl has a Working With Children Check (WWCC), others in the Department will be working with the schools and parents to collect this data, most of which will be online.

The survey data will be collected and stored per NSW Department of Education guidelines, including those relating to both data security and child protection. The NSW Department of Education has reviewed and approved the survey instruments for this research. Formal consent will be necessary from each participant.

Yours sincerely,

Production Note:
Signature removed
prior to publication.

Lisa Harrington
Executive Director, Business Enablement, School Infrastructure NSW

SCHOOL INFRASTRUCTURE NSW
259 George Street Sydney NSW 2000   PO Box 33 Sydney NSW 2001   T 9273 9200
education.nsw.gov.au

Appendix 3. Execution of X-means clustering and Gradient Boosted Trees in RapidMiner

The execution of X-means cluster analysis in RapidMiner for this research is shown in Figure A.1. Subsequent sub-paragraphs explain the individual operators' purpose and decisions taken at each step.

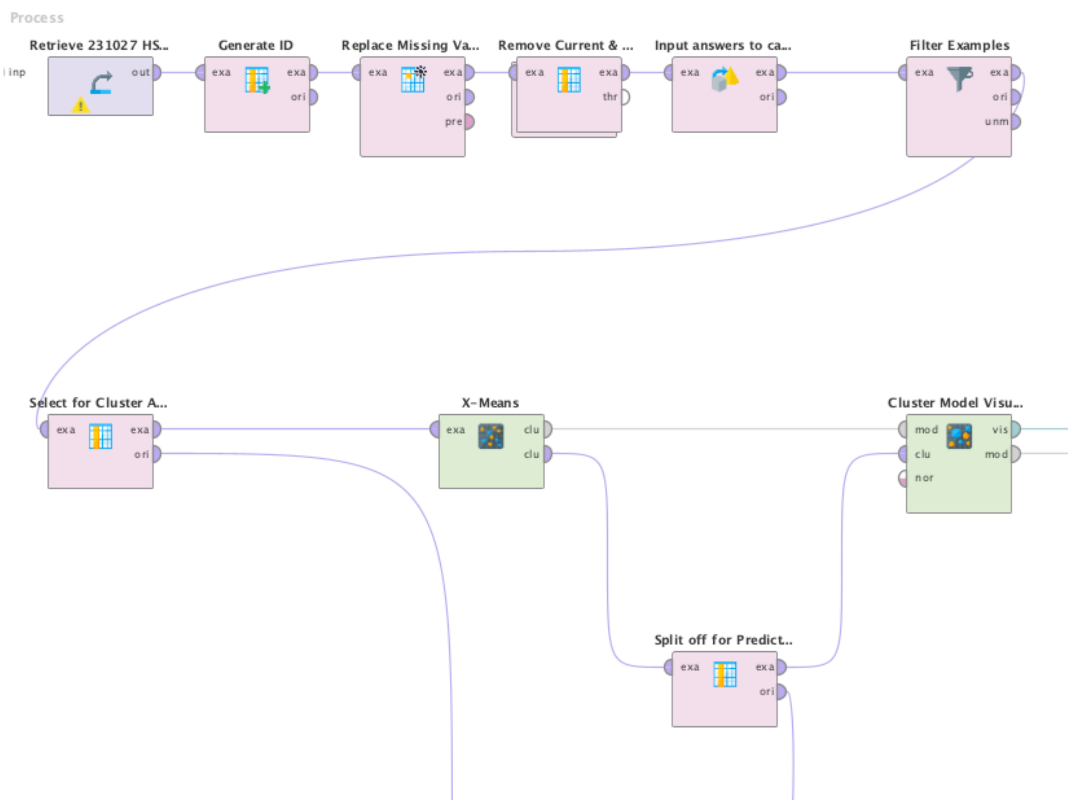

Figure A.1. A screenshot of X-means cluster analysis in RapidMiner

Each of the following sub-paragraphs explains the function of each object shown in Figure A.1.The description starts from the second object in the top left of the diagram; the first operator simply retrieves the data containing every survey response for analysis.

Generate ID. The X-Means Cluster analysis is only applied to the outcomes construct from the survey. As such, there is a need to separate outcomes and inputs. However, the predictive analysis component requires access to both inputs and outcomes for each respondent. As such, the data must be rejoined. The Generate ID operator generates a unique nominal identifier for each respondent, which is carried forward.

Replace Missing Values. X-means clustering is not able to process datasets with missing values. Missing values were replaced with averages (numerical) or most common (nominal) values.

Respondent IDs with missing values were logged to enable identification and exclusion as required.

Remove Current and Preferred. Respondents nominate a current and preferred experience when addressing non-infrastructure inputs. When importing the original CSV files, an additional attribute was created by calculating the difference between these two responses. As such, the two original responses were no longer required and were removed. This step is required for the (later) predictive analysis.

Input answers to categories. This operator ensures that the input answers are nominal in type. This step is required for the (later) predictive analysis.

Filter Examples. This operator enables the process (X-Means and Gradient Boosted Tree) to receive either all responses or responses from groups, such as individual schools.

Select for Cluster Analysis. This operator selects those attributes to be used in clustering. This operator is used to exclude all but the Protective and Risk Factor responses. The ID is also passed through to ensure the integrity of the later combined data set.

X-Means. This operator controls and conducts the X-means clustering process. Decisions made at this step include the minimum number of clusters produced and the means of defining the initial centre of each cluster.

Cluster Model Visualisation. This operator increases the interpretability of the X-means analysis. This operator produces four visualisations: a graphical description of the clusters, a heat map, a centroid chart, and a centroid table. The latter two visualisations are most useful as they show the difference in the clusters across the entire attribute range. Cluster Model Visualisation is the last step in the X-means cluster analysis. The dataset from this analysis is forwarded to the Gradient Boosted Tree part of the process, which is explained next.

Appendix 4. Predictive Analysis in RapidMiner

The execution of GBT analysis in RapidMiner for this research is shown in Figure A.2.

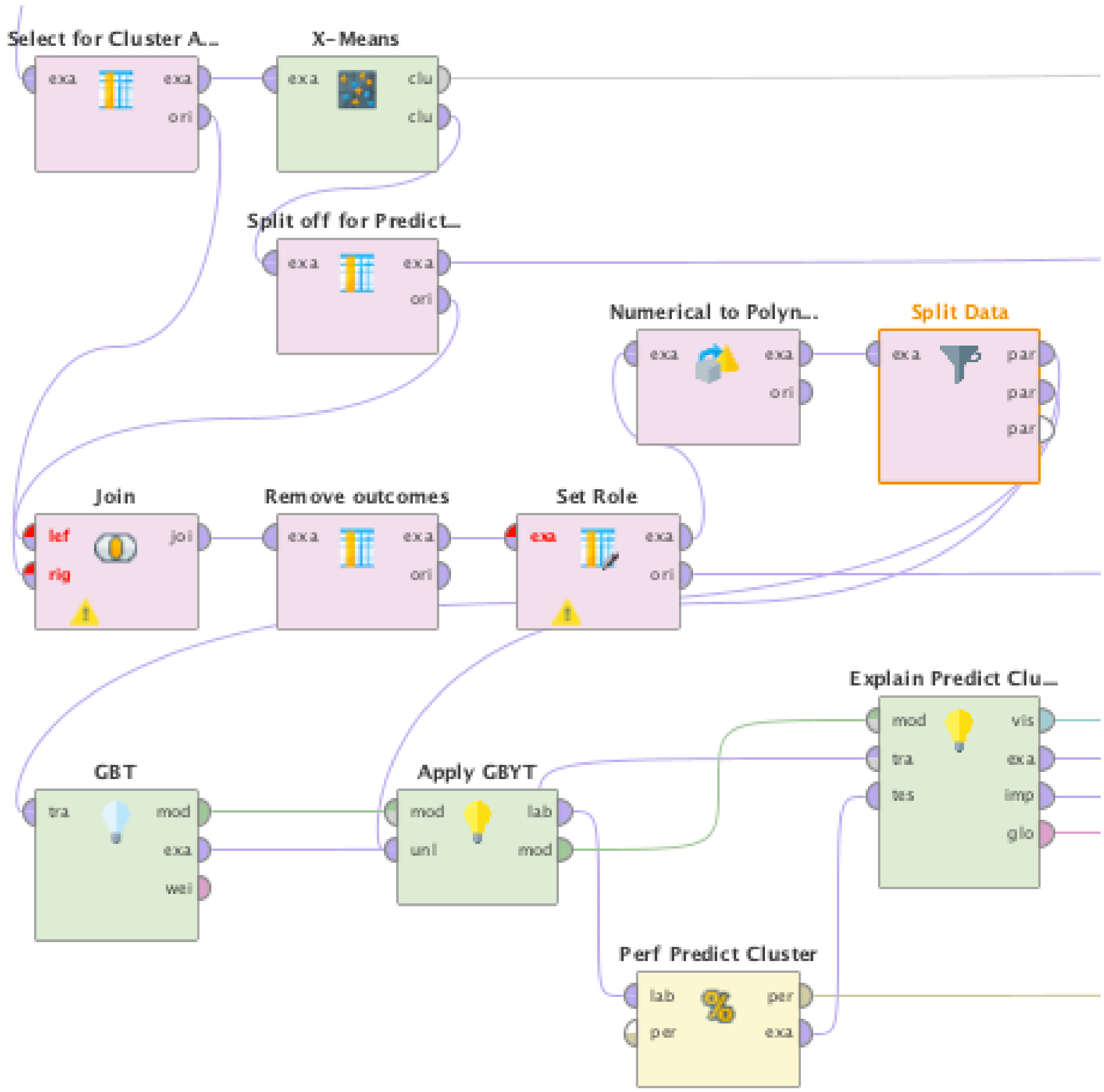

Figure A.2. A screenshot of GBT analysis in RapidMiner

Each of the following sub-paragraphs explains the function of each object shown in Figure A.2. The description starts with the left-most object, the Join operator.

Join. The X-Means process required that the outcomes be separated from the inputs. The Join operator brings together the outcomes (Protective and Risk factors) and inputs (Safety, Interpersonal Relationships, Teaching and Learning, and Institutional Environment). The outcomes dataset is required because it contains the cluster attribute.

Remove outcomes. This operator removes the outcome responses from the outcome dataset. The effect of this is to leave the Cluster attribute as the only outcome attribute. Without this step, the outcome responses would be used in predicting the outcome clusters.

Set Role. Each model has a target variable. This is the attribute that the model tries to predict. This step selects the target variable. In the case of this research, the cluster is the target variable.

Numerical to Polynomial. Decision trees are a categorisation technique, as opposed to a numerical technique. As such, all numerical data must be converted to polynomials. This conversion is critical because many questions ask the respondents to answer between one and five, corresponding to a Likert Scale word picture.

Split Data. The model requires a training dataset and a test (or application) dataset. Two decisions were made and executed at this point. The first is the relative size of the two data sets; the second is the means of populating these data sets. For this research, both sets (training and test) are the same size, each containing 50% of the data set. Random sampling with a local seed populates each set. This approach requires that each sample (a complete response from one participant) is given a unique, sequential, numeric identifier. A random selection process then allocates (using the identifier) the response to either the training or test set. Discussed further in the Results chapter are the practical limitations imposed by splitting the data. Specifically, learning techniques are usually more successful when using more extensive datasets. Unfortunately, splitting the dataset halves the data available to the model and has ramifications when considering the level at which predictions are made.

Gradient Boosted Tree (GBT). From the Split Data operator, there are two separate data paths. One is to the Application operator; the other is to the GBT operator. This operator contains the algorithm that learns from the data to produce the model, the GBT. As explained in more detail in the Results chapter, the *algorithm parameters* are set in this step. These parameters define how the nodes split into branches and leaves and the tree's overall size (or depth). As with most analytical decisions, there is a trade-off between complexity, efficiency, and interpretability.

Apply GBT. Two inputs feed the Apply GBT operator: the model (from the GBT) operator and the test set from the Split Data operator. The model is provided with the test data set in this step, and the Target Variable is withheld. The model is then run, with the predictions passed (along with the Cluster) to the Performance operator.

Perf Predict Cluster. The model's objective is to predict the values of the Target Variable using an unseen (test) data set. In the case of this research, the Cluster is the target variable. The predicted (from the model) and actual (from the test data) values of the Cluster attribute are then passed to the Perf Predict Cluster operator. This operator then analyses the efficiency and effectiveness of the model. This operator also renders the visual representation of the decision tree. This operator produces a dataset containing the results of every prediction and an assessment of the accuracy of the GBT Model, defined as the average of the precision and recall. The accompanying diagram (Figure A.3) illustrates the differences between the complementary measures of accuracy and precision.

	true cluster_0	true cluster_1	class precision
pred. cluster_0	1132	359	75.92%
pred. cluster_1	423	732	63.38%
class recall	72.80%	67.09%	

Figure A.3. A confusion matrix executed by RapidMiner

Explain Predict Cluster. The final operator in this process is the explanatory operator. The purpose of this operator is to visualise the GBT model and to calculate and tabulate the contribution (weights) of each of the inputs in predicting the cluster to which each respondent belongs.

The process as executed in RapidMiner is executed each time there is a different scenario to be examined. In all cases, the results are written to one dataset for each scenario. The results were initially analysed in RapidMiner before being exported to MS Excel for further analysis and visualisation.

Appendix 5. School Climate Misfit: Descriptive statistics

Shown in Figure A.4. are the descriptive statistics for the School Climate misfit. The mean for these responses is shown at Figure 4.1. For each attribute the range (thin capped lines) is four, indicating that for each of these attributes there are students that prefer the current situation, and students who prefer a very different situation.

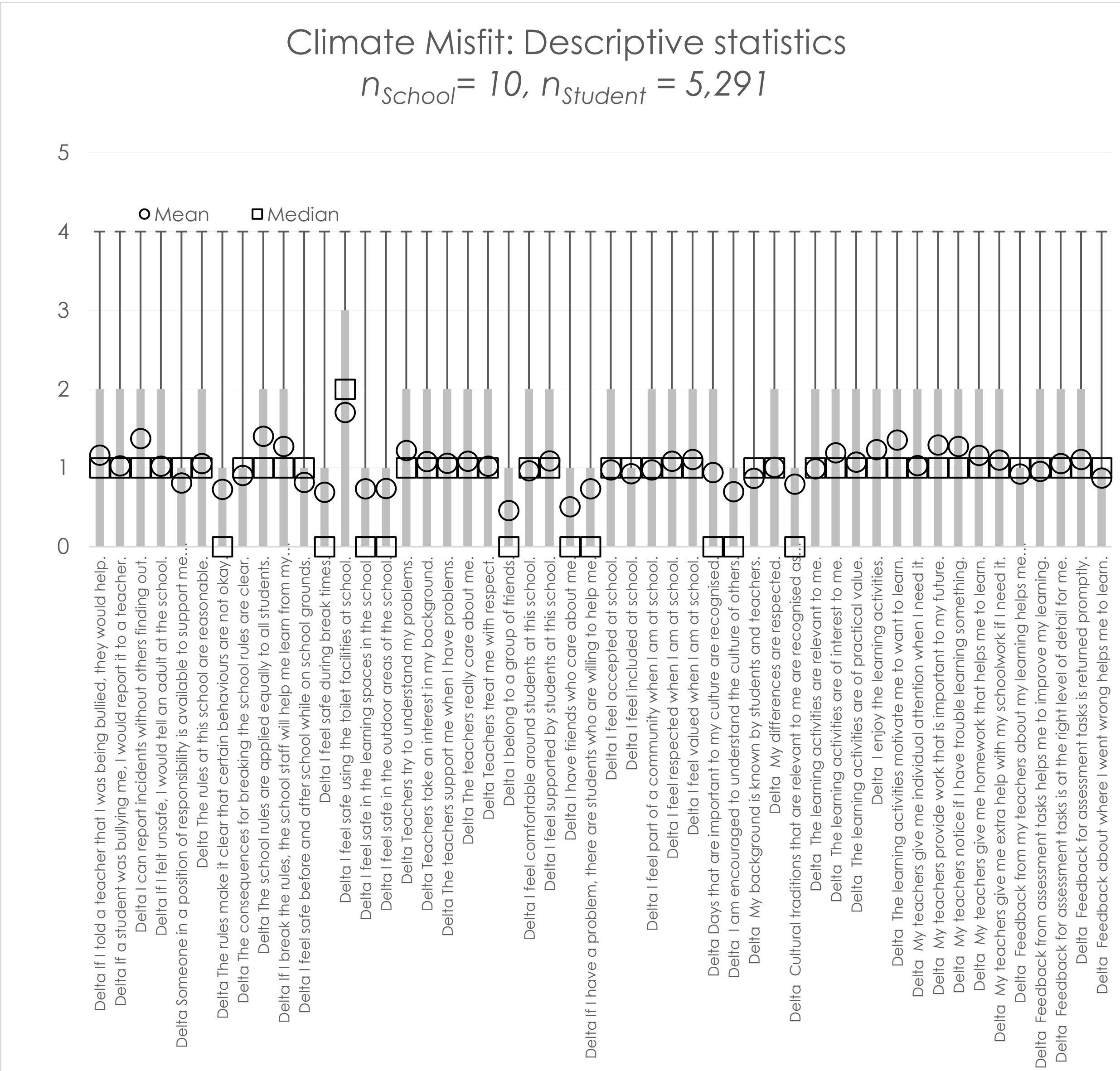

Figure A.4. Climate misfit descriptive statistics

Appendix 6. Institutional Environment: Descriptive statistics

Figure A.5 shows the response data for the Secondary school Institutional Environment. The Institutional Environment responses are scored from one to five, utilising a five-point Likert scale, spanning strong disagreement to strong agreement respectively. Students responded with their current view of the Institutional Environment only, and so no calculation of difference (misfit) required.

In all cases, the responses span the whole scale. Responses were provided using a five-point Likert scale, and the midpoint or neutral response is three. As can be seen in Figure A.5 for most of the elements, the responses are skewed toward the *negative*, that is, students disagreeing (to some degree) with the offered statement. Those areas where most of the students disagreed with (to some extent) with the positive statement of experience, specifically maintenance of furniture and fittings, cleanliness and tidiness of the school, temperature, smell, intrusive noise, comfortable furniture, and shelter from the elements.

While no elements had a median exceeding three, there are some elements where the mean response is skewed toward the positive. These elements are maintenance of school grounds, classroom lighting, line of sight to information in classrooms, suitability of classrooms for different subjects, availability of space for students wanting to be active and noisy, availability of space for students to socialise, and the selection of outdoor activities.

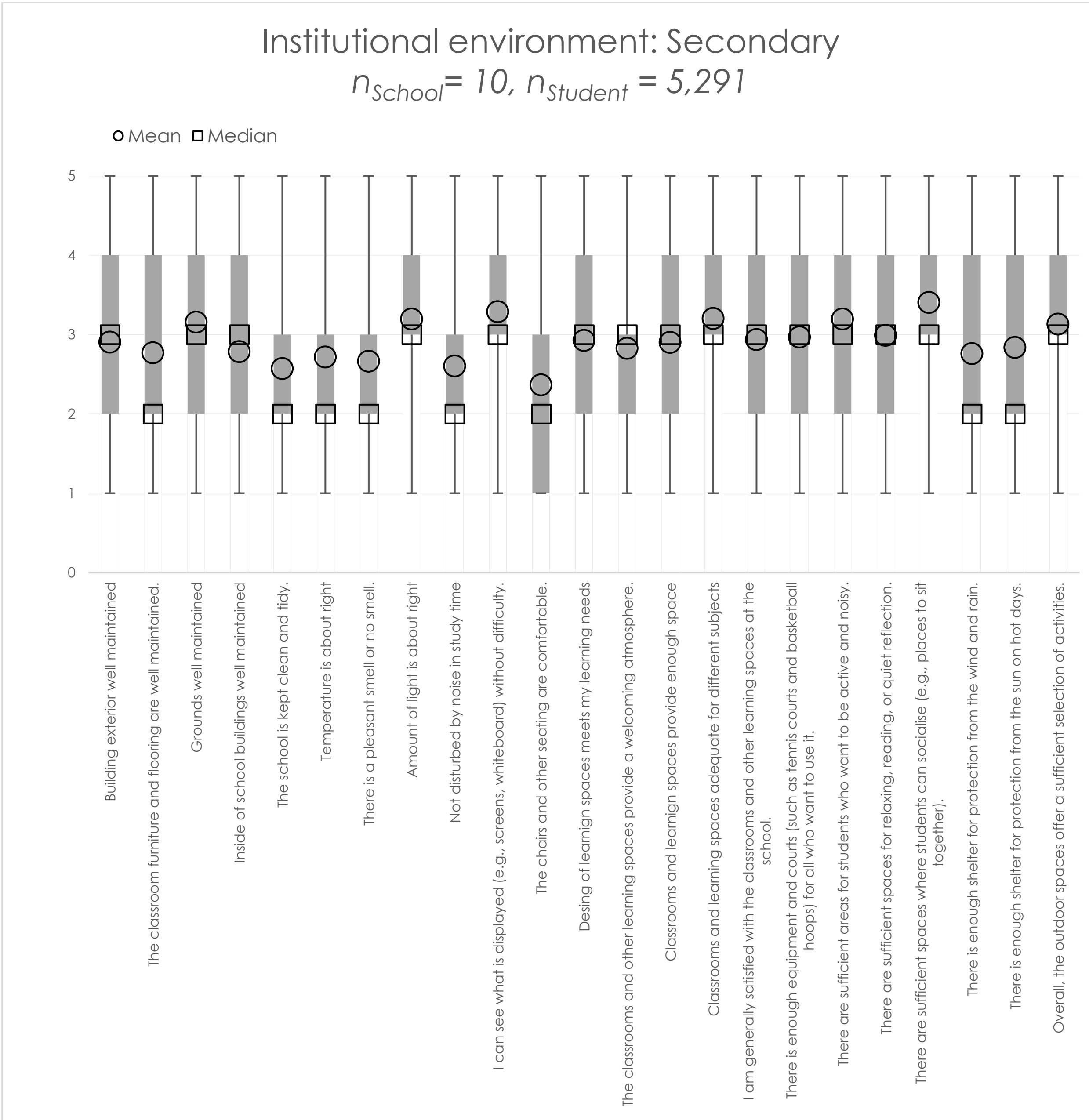

Figure A.5. Institutional environment descriptive statistics

Appendix 7. School specific clustering

This Appendix contains charts showing the outcome clusters for each school.

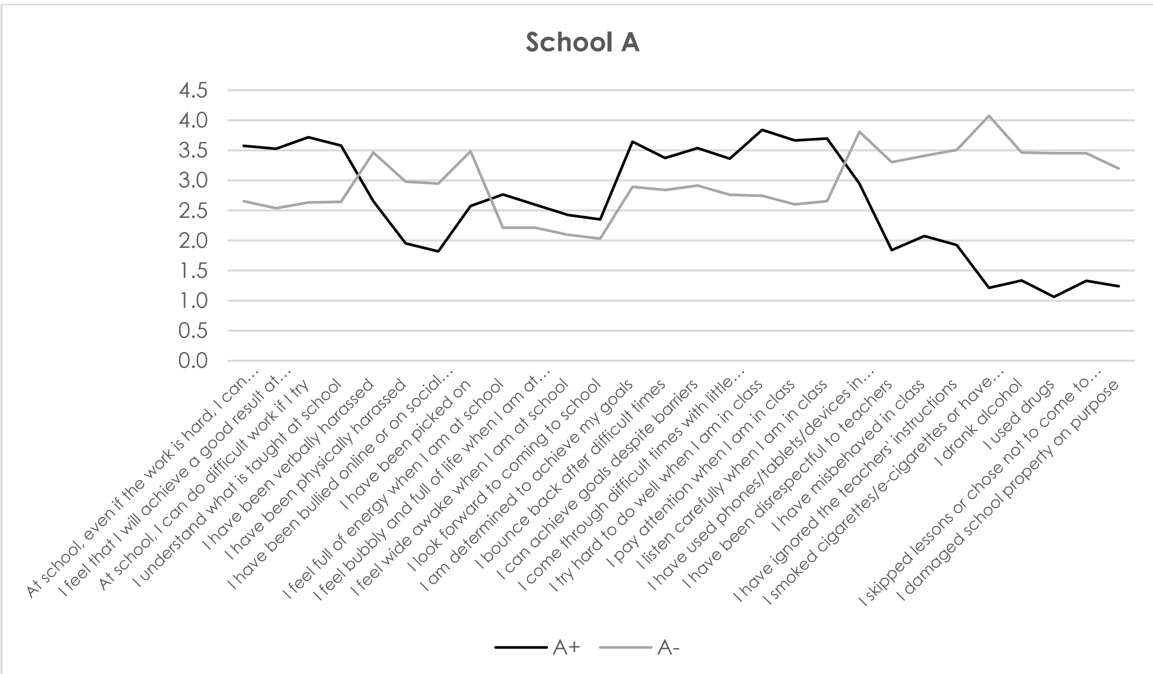

Figure A. 6. School A: Outcome clusters

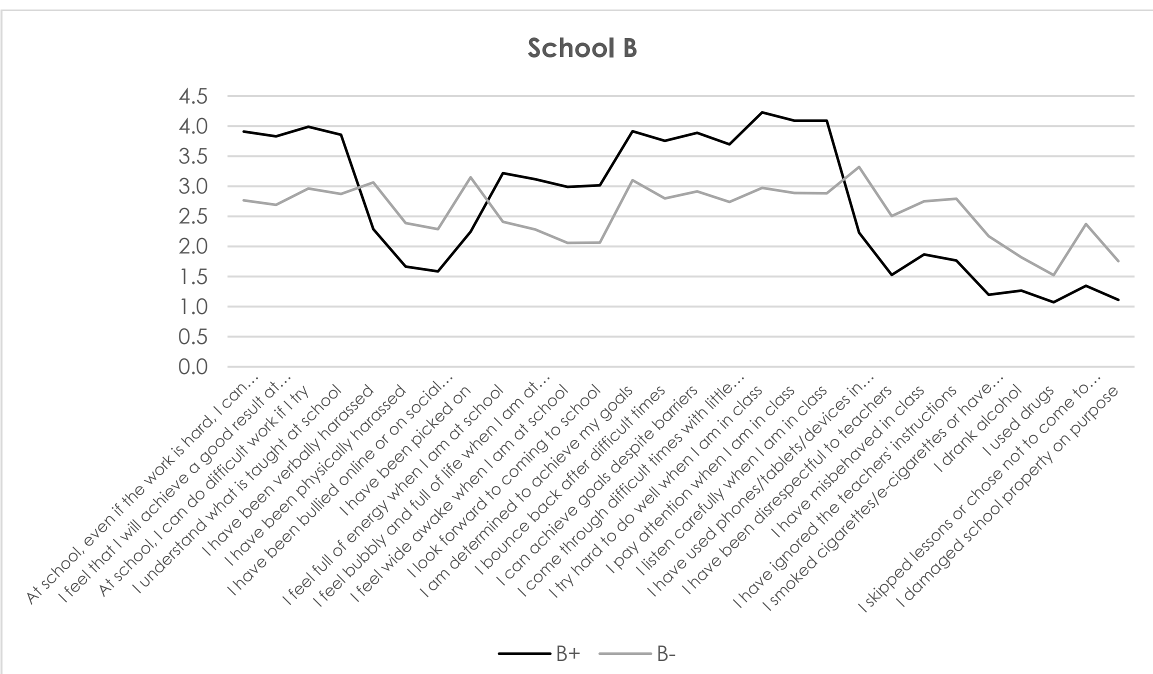

Figure A. 7. School B: Outcome clusters

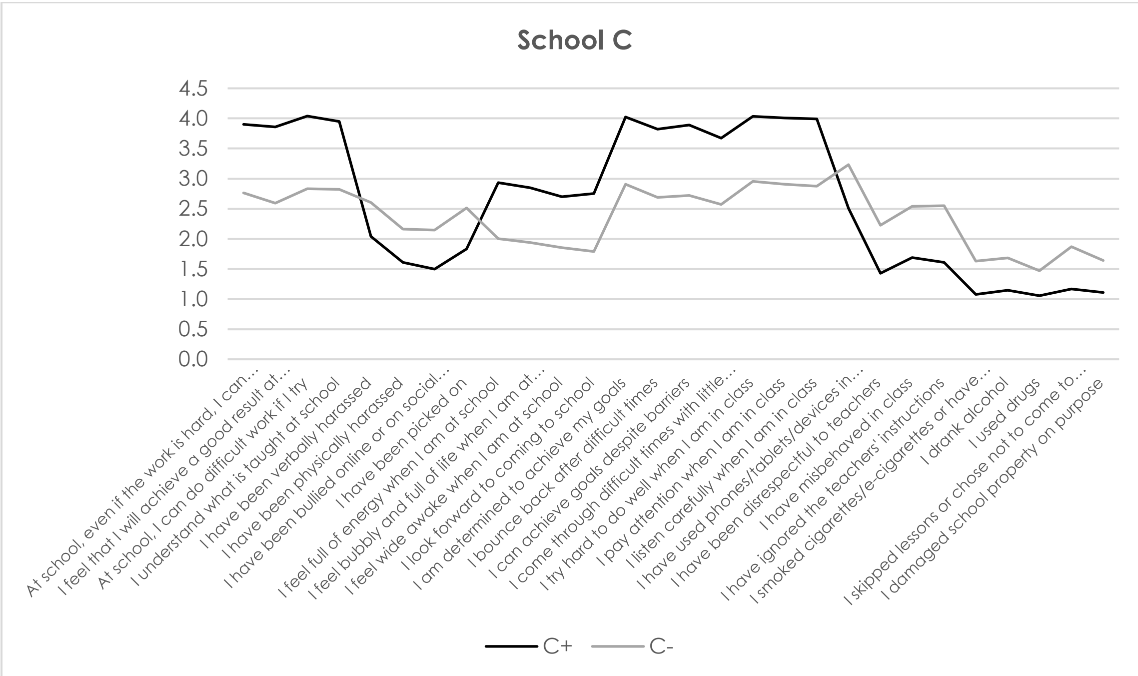

Figure A. 8. School C: Outcome clusters

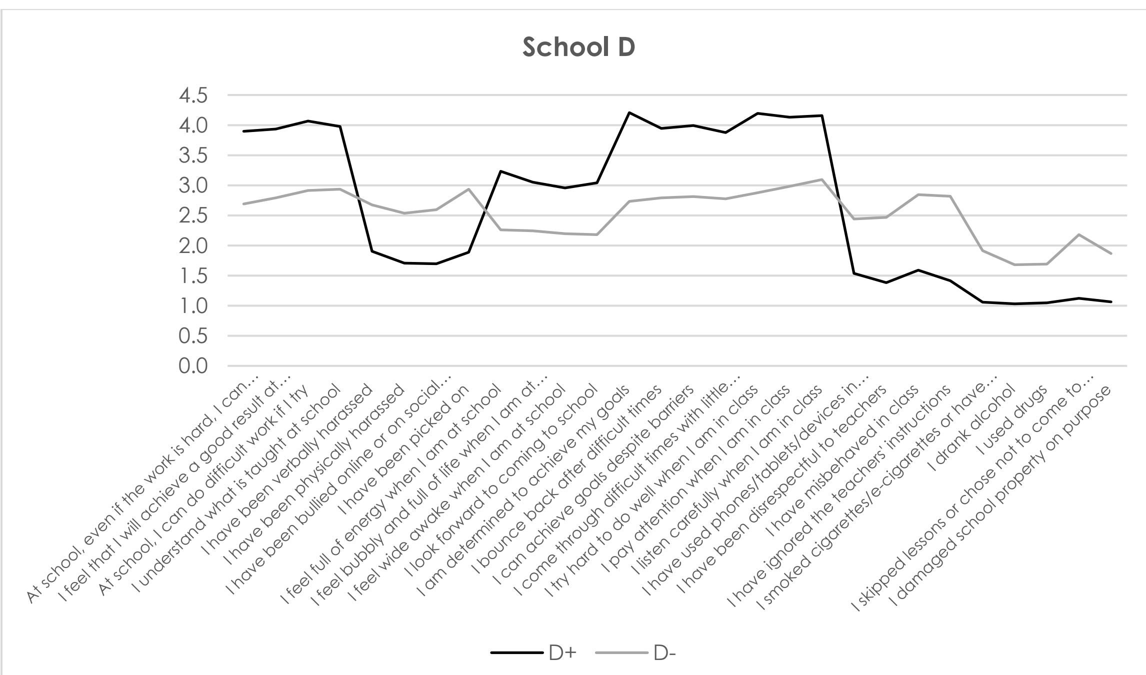

Figure A. 9. School D: Outcome clusters

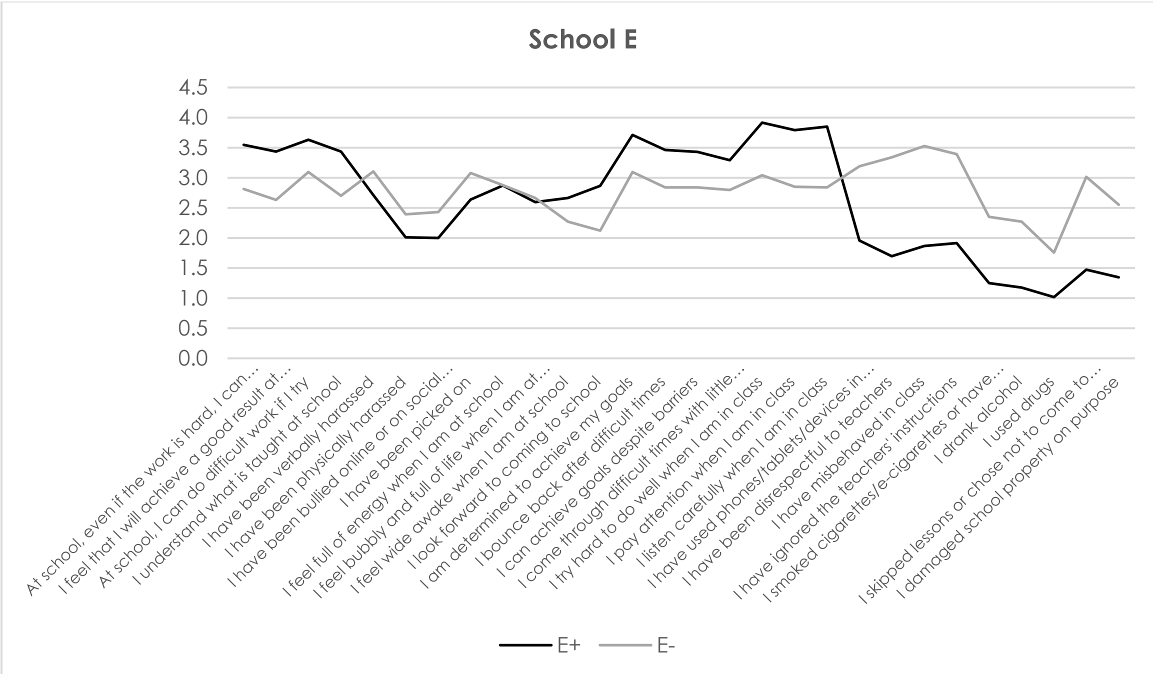

Figure A. 10. School E: Outcome clusters

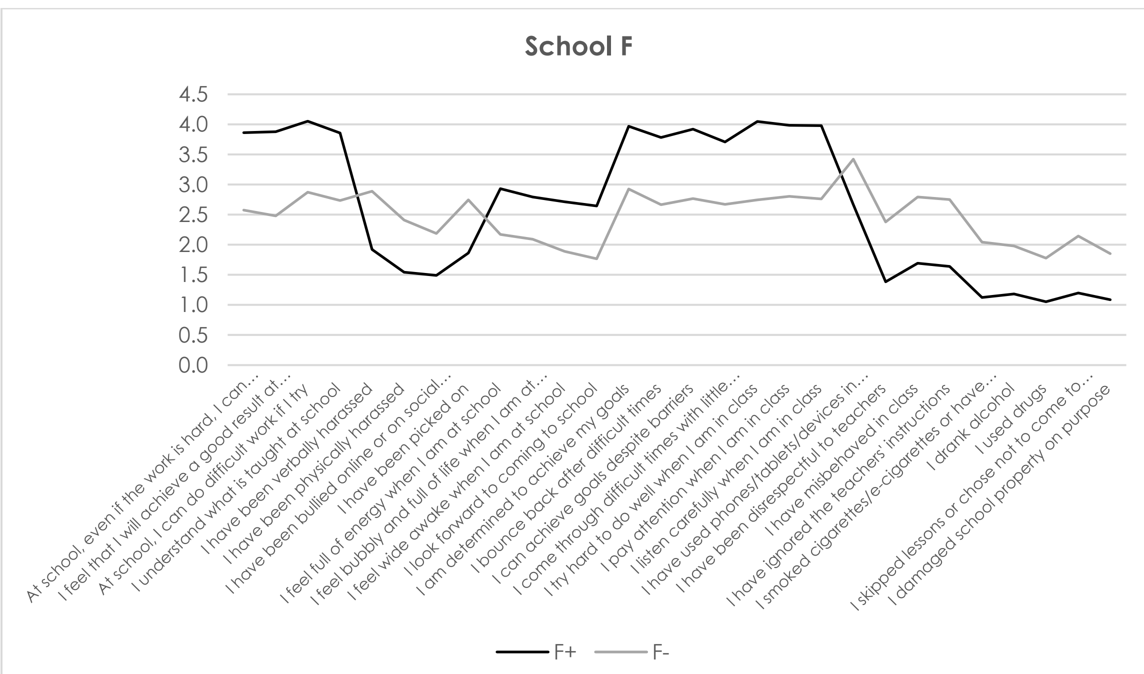

Figure A. 11. School F: Outcome clusters

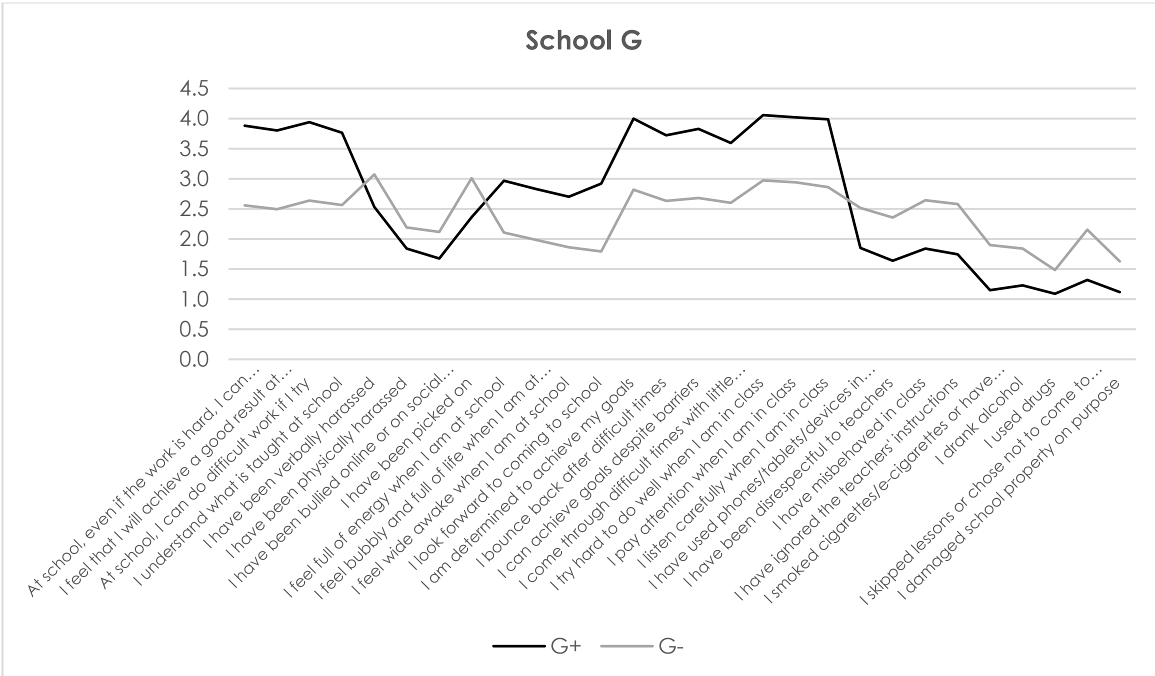

Figure A. 12. School G: Outcome clusters

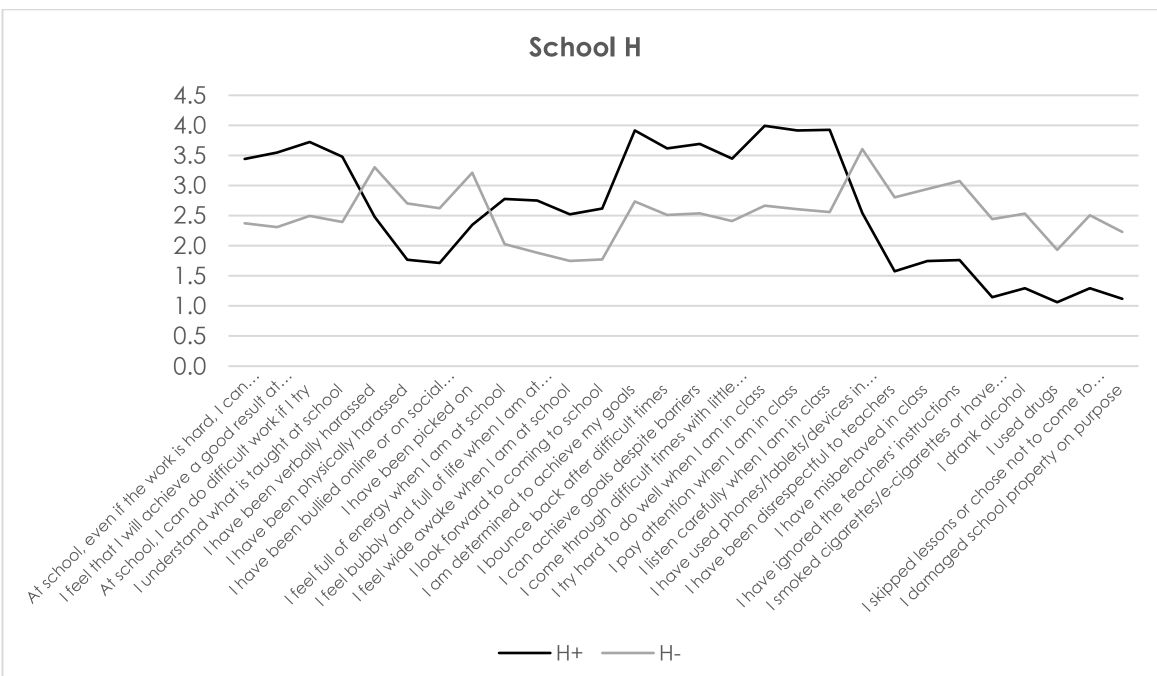

Figure A 13. School H: Outcome clusters

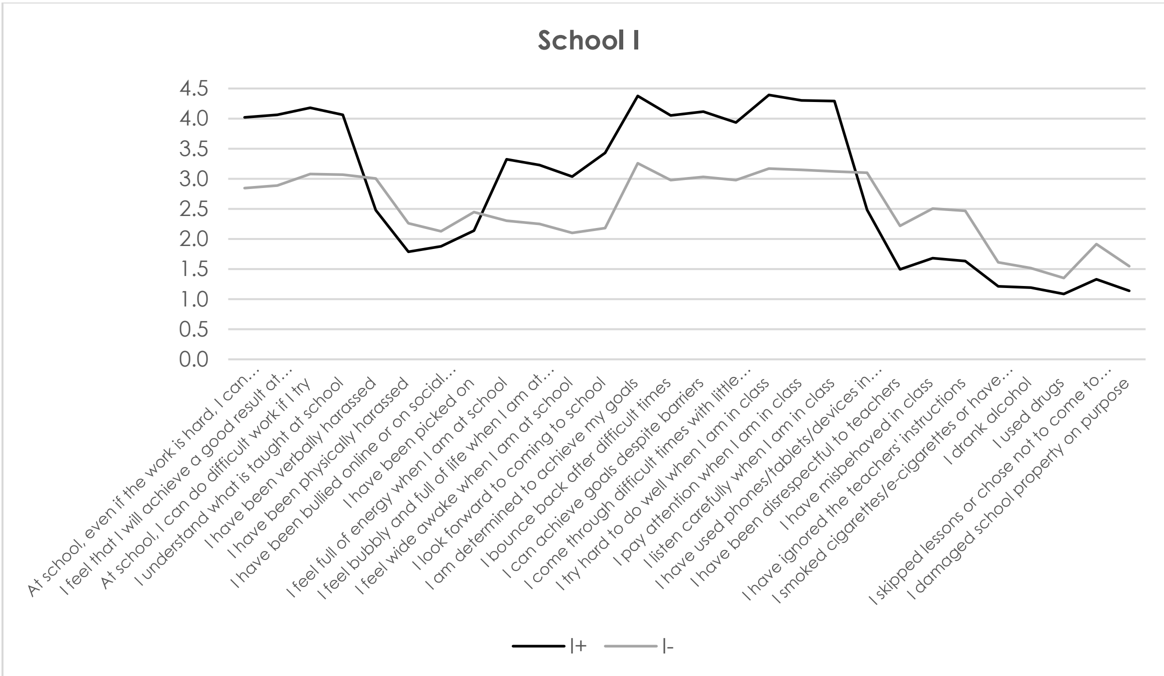

Figure A. 14. School I: Outcome clusters

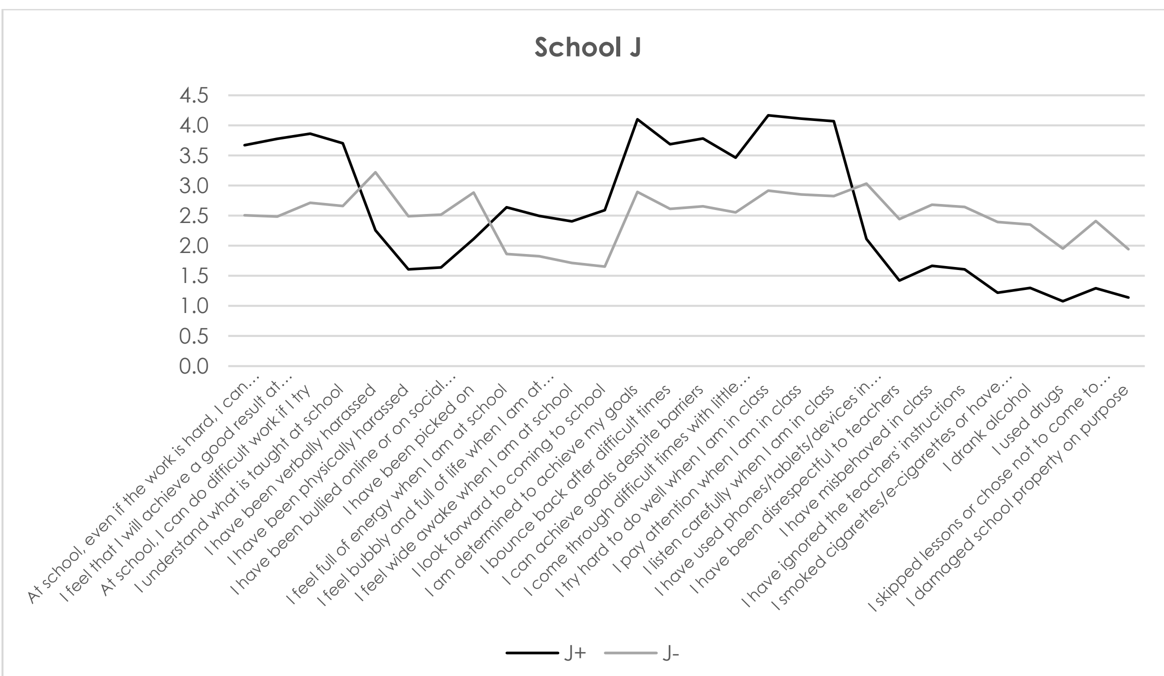

Figure A. 15. School J: outcome clusters

Appendix 8. Gradient boosted tree predictive weights

Pillar	Construct	Predictive attribute	School A	School B	School C	School D	School E	School F	School G	School H	School I	School J
Academic	Engaging environment	Delta I enjoy the learning activities	0.053	0.142	0.097	0.086	0.078	0.326	0.041	0.259	0.063	0.098
		Delta The learning activities are of interest to me	0.083	0.144	0.144	0.060	0.116	0.095	0.316	0.108	0.130	0.099
		Delta The learning activities are of practical value	0.200	0.366	0.193	0.154	0.225	0.132	0.100	0.094	0.066	0.141
		Delta The learning activities are relevant to me	0.054	0.056	0.089	0.052	0.099	0.371	0.101	0.158	0.137	0.224
		Delta The learning activities motivate me to want to learn	0.096	0.052	0.134	0.052	0.169	0.251	0.166	0.142	0.125	0.111
	Support for learning	Delta My teachers give me extra help with my schoolwork if I need it	0.198	0.171	0.207	0.203	0.034	0.258	0.128	0.472	0.043	0.259
		Delta My teachers give me homework that helps me to learn	0.142	0.062	0.163	0.112	0.287	0.208	0.116	0.083	0.101	0.109
		Delta My teachers give me individual attention when I need it	0.302	0.038	0.119	0.058	0.130	0.132	0.069	0.317	0.328	0.092
		Delta My teachers notice if I have trouble learning something	0.128	0.047	0.115	0.114	0.254	0.163	0.065	0.260	0.082	0.218
		Delta My teachers provide work that is important to my future	0.073	0.166	0.061	0.134	0.112	0.243	0.197	0.094	0.226	0.091
	Teacher feedback	Delta Feedback about where I went wrong helps me to learn	0.052	0.069	0.116	0.037	0.099	0.187	0.097	0.051	0.222	0.136
		Delta Feedback for assessment tasks is at the right level of detail for me	0.035	0.052	0.327	0.086	0.087	0.241	0.035	0.119	0.170	0.224
		Delta Feedback for assessment tasks is returned promptly	0.065	0.090	0.072	0.099	0.042	0.077	0.133	0.198	0.302	0.149
		Delta Feedback from assessment tasks helps me to improve my learning	0.080	0.049	0.084	0.094	0.060	0.073	0.066	0.064	0.078	0.057
		Delta Feedback from my teachers about learning helps me understand how to improve	0.056	0.064	0.108	0.096	0.082	0.057	0.240	0.117	0.167	0.069
Community	Affirming diversity	Delta I am encouraged to understand the culture of others	0.265	0.235	0.110	0.050	0.139	0.074	0.088	0.094	0.132	0.156
		Delta Cultural traditions that are relevant to me are recognised as important	0.104	0.104	0.063	0.043	0.110	0.094	0.051	0.092	0.268	0.081
		Delta Days that are important to my culture are recognised	0.061	0.145	0.115	0.431	0.126	0.119	0.163	0.100	0.093	0.112
		Delta My background is known by students and teachers	0.068	0.059	0.341	0.208	0.137	0.573	0.080	0.137	0.336	0.270
		Delta My differences are respected	0.138	0.074	0.150	0.118	0.027	0.149	0.045	0.046	0.221	0.119
	Peer connectedness	Delta I belong to a group of friends	0.325	0.047	0.090	0.074	0.062	0.307	0.149	0.045	0.211	0.079
		Delta I feel comfortable around students at this school	0.097	0.035	0.174	0.219	0.029	0.165	0.047	0.070	0.115	0.074
		Delta I feel supported by students at this school	0.107	0.141	0.155	0.060	0.070	0.136	0.156	0.055	0.107	0.069
		Delta I have friends who care about me	0.057	0.068	0.127	0.135	0.102	0.111	0.089	0.057	0.267	0.622
		Delta If I have a problem, there are students who are willing to help me	0.067	0.086	0.095	0.083	0.027	0.184	0.041	0.060	0.071	0.055
	School connectedness	Delta I feel accepted at school	0.098	0.048	0.076	0.138	0.156	0.171	0.175	0.078	0.143	0.091
		Delta I feel included at school	0.257	0.071	0.170	0.135	0.082	0.156	0.079	0.085	0.082	0.287
		Delta I feel part of a community when I am at school	0.555	0.127	0.133	0.200	0.164	0.065	0.065	0.106	0.077	0.328
		Delta I feel respected when I am at school	0.096	0.131	0.302	0.202	0.043	0.136	0.145	0.099	0.144	0.214
		Delta I feel valued when I am at school	0.086	0.052	0.219	0.078	0.296	0.065	0.102	0.062	0.132	0.087
	Teacher support	Delta Teachers take an interest in my background	0.042	0.318	0.139	0.080	0.116	0.070	0.589	0.064	0.292	0.108
		Delta Teachers treat me with respect	0.105	0.092	0.102	0.038	0.029	0.146	0.095	0.075	0.191	0.107
		Delta Teachers try to understand my problems	0.043	0.270	0.103	0.097	0.093	0.157	0.052	0.028	0.375	0.069
		Delta The teachers really care about me	0.594	0.059	0.236	0.041	0.056	0.608	0.120	0.147	0.149	0.175
		Delta The teachers support me when I have problems	0.050	0.062	0.234	0.128	0.034	0.114	0.271	0.064	0.183	0.089
Institutional environment	Ambient environment	Amount of light is about right	0.291	0.496	0.174	0.138	0.081	0.749	0.822	0.068	0.048	0.365
		I can see what is displayed (e g , screens, whiteboard) without difficulty	0.206	0.287	0.059	0.787	0.770	0.439	0.044	0.296	0.090	0.431
		Not disturbed by noise in study time	0.074	0.052	0.398	0.056	0.112	0.192	0.076	0.073	0.076	0.145
		Temperature is about right	0.726	0.039	0.143	0.230	0.265	0.429	0.076	1.000	0.070	0.088
		There is a pleasant smell or no smell	0.101	1.000	0.082	0.126	0.448	0.069	0.164	0.147	0.248	0.148
	Condition, maintenance and upkeep	Building exterior well maintained	0.413	0.080	0.170	0.759	0.891	0.173	0.426	0.103	0.101	0.161
		Grounds well maintained	0.304	0.070	0.272	0.265	0.083	0.273	0.070	0.470	0.780	0.280
		Inside of school buildings well maintained	0.141	0.066	0.072	0.218	0.250	0.123	0.093	0.041	0.396	0.181
		The classroom furniture and flooring are well maintained	1.000	0.049	0.117	0.066	0.157	0.196	0.160	0.097	0.175	0.156
		The school is kept clean and tidy	0.087	0.049	0.055	0.312	0.421	0.582	0.345	0.542	1.000	0.358
	Design indoor	Classrooms and learning spaces adequate for different subjects	0.057	0.458	0.143	0.056	1.000	0.509	0.084	0.074	0.072	1.000
		Classrooms and learning spaces provide enough space	0.402	0.154	0.588	0.824	0.033	0.528	0.241	0.382	0.974	0.484
		Design of learning spaces meets my learning needs	0.089	0.182	1.000	0.135	0.032	0.192	0.330	0.091	0.820	0.604
		I am generally satisfied with the classrooms and other learning spaces at the school	0.055	0.040	0.094	0.102	0.043	1.000	0.132	0.333	0.723	0.131
		The classrooms and other learning spaces provide a welcoming atmosphere	0.082	0.138	0.357	0.047	0.049	0.087	1.000	0.077	0.185	0.090
	Design outdoor	Overall, the outdoor spaces offer a sufficient selection of activities	0.070	0.075	0.100	0.058	0.097	0.320	0.284	0.363	0.103	0.065
		There are sufficient areas for students who want to be active and noisy	0.566	0.134	0.133	0.420	0.142	0.738	0.119	0.352	0.165	0.140
		There are sufficient spaces for relaxing, reading, or quiet reflection	0.171	0.035	0.112	0.056	0.036	0.155	0.097	0.381	0.087	0.065
		There are sufficient spaces where students can socialise	0.235	0.310	0.100	0.122	0.021	0.227	0.080	0.440	0.715	0.864
		There is enough equipment and courts for all who want to use it	0.040	0.178	0.076	0.056	0.039	0.163	0.194	0.047	0.302	0.094
		There is enough shelter for protection from the sun on hot days	0.092	0.082	0.113	1.000	0.711	0.059	0.058	0.100	0.234	0.065
		There is enough shelter for protection from the wind and rain	0.231	0.085	0.081	0.435	0.198	0.116	0.042	0.040	0.109	0.118
	Ergonomics	The chairs and other seating are comfortable	0.053	0.040	0.160	0.045	0.155	0.174	0.063	0.059	0.380	0.098
Safety	Reporting and seeking help	Delta I can report incidents without others finding out	0.294	0.149	0.079	0.093	0.069	0.142	0.039	0.138	0.166	0.277
		Delta If a student was bullying me, I would report it to a teacher	0.290	0.047	0.116	0.230	0.758	0.220	0.118	0.206	0.119	0.116
		Delta If I felt unsafe, I would tell an adult at the school	0.046	0.167	0.074	0.036	0.116	0.085	0.151	0.182	0.138	0.142
		Delta If I told a teacher that I was being bullied, they would help	0.416	0.142	0.068	0.164	0.046	0.111	0.091	0.128	0.056	0.709
		Delta Someone in a position of responsibility is available to support me if I need it	0.239	0.056	0.241	0.071	0.144	0.206	0.096	0.229	0.107	0.170
	Rule Clarity	Delta If I break the rules, the school staff will help me learn from my mistake	0.059	0.101	0.092	0.080	0.094	0.287	0.113	0.170	0.079	0.131
		Delta The consequences for breaking the school rules are clear	0.038	0.030	0.185	0.111	0.060	0.333	0.163	0.046	0.429	0.089
		Delta The rules at this school are reasonable	0.147	0.131	0.208	0.086	0.138	0.102	0.046	0.061	0.168	0.129
		Delta The rules make it clear that certain behaviours are not okay	0.109	0.090	0.084	0.053	0.045	0.109	0.094	0.076	0.267	0.071
		Delta The school rules are applied equally to all students	0.206	0.070	0.162	0.068	0.024	0.155	0.065	0.120	0.127	0.082
	Safe environment	Delta I feel safe before and after school while on school grounds	0.267	0.060	0.074	0.076	0.101	0.193	0.111	0.083	0.057	0.054
		Delta I feel safe during break times	0.046	0.096	0.221	0.097	0.123	0.083	0.084	0.042	0.142	0.115
		Delta I feel safe in the learning spaces in the school	0.098	0.090	0.280	0.050	0.070	0.112	0.242	0.259	0.176	0.168
		Delta I feel safe in the outdoor areas of the school	0.048	0.161	0.094	0.073	0.078	0.085	0.044	0.133	0.062	0.107
		Delta I feel safe using the toilet facilities at school	0.056	0.061	0.210	0.109	0.036	0.063	0.034	0.088	0.110	0.138

Figure A. 16. Predictive weights (Red<10th percentile, Green>90th percentile)