BUILDING A CLIMATE FOR CREATIVITY:

A THEORY OF ACTION TO IMPROVE U.S. ELEMENTARY SCHOOLS

James P. Patsalides

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

From

Prescott College

In

Sustainability Education

May 2015

Jacqueline Kelleher, Ph.D.	*Kara Sammet, Ph.D.*
Committee Member	Committee Member
Noël Cox Caniglia, Ph.D.	*Denise Mitten, Ph.D.*
Committee Chair	External Reader

James P. Patsalides
55 New Street
Manchester, Connecticut 06040

jpatsalides@prescott.edu

TABLE OF CONTENTS

Tables

Figures

Abstract

In the context of a rapidly changing world, higher order thinking skills are necessary for sustainability of U.S. society. Beginning with the premise that U.S. public schools are charged with the constitutional and moral duty of growing children into informed and educated citizens, prepared to thrive in the world of work *and* to participate in democratic processes; and, that higher order thinking is a core part of that mission, this study examined children's perspectives on school climate and the environment for the teaching and learning of higher order thinking in twenty five public elementary schools in an urban Connecticut school district. This integrated program of research used an exploratory sequential/concurrent mixed methods design to construct a pair of new psychometric instruments to measure student attitudes toward school climate and the environment for teaching and learning higher order thinking in a public elementary school. The intended uses and interpretations of the scores reported by the Climate4Creativity Elementary (C4C/SPE) and Middle School (C4C/SPM) Student Perspectives measurement instruments, were validated to professional standards. The study concluded that these instruments have utility for public elementary schools, particularly in identifying areas of focus and in the management of strategic and tactical school improvement work as part of a wider program of transformation in a school. Cronbach's Alpha reliability scores in excess of 0.90 were reported for all measures. This study supported the core idea that safer schools with stronger, more caring communities provide individual students with better learning environments, and that general learning and the learning of creativity are intrinsically linked in the minds of students in public elementary schools, even though these students may not always name these components as such. Attitudes toward the environment for higher order thinking tend to deteriorate from the early grades to middle school grades, implying both raised

expectations, and an increase in variability in the data due to more and more variety in classroom settings and teacher practices. Examination of reported bullying experience shows bullying victimization to be a powerful, pervasive determinant of school climate and feelings of safety and community in all grades, but, bullying victimization tends not to penetrate into perceptions of the classroom learning environment to the same degree. By exploring school safety, community, and the structure of the learning environment required for the teaching and learning of higher order thinking in a public elementary school, this work begins the creation of a framework to enable school leaders to make significant, transformational, strategic change in their schools.

Keywords: school climate, learning environment, creativity, Rasch measurement, school improvement.

Acknowledgements

My journey from professional manager to professional educator has been the most challenging and rewarding transformation of my life. There are so many human beings who have helped me on this course and enabled me to find my *mojo* as a researcher and as a newly minted educational activist. I am particularly thankful for the circle of incredible, intelligent, dedicated, beautiful women who have surrounded and nurtured me these past few years.

First, my thanks to Phi Delta Kappa, the honor society for teachers, for providing $1,510 funding to offset the costs of completing this research. Your mission to improve education by supporting professional teachers and professional research on teaching and learning resonates with me on so many levels. Thank you for everything you do in the world.

To my committee chair, feminist ethnographer, woman rancher, advisor, and friend, Dr. Noël Cox Caniglia. Noël: for our monthly phone calls these past two years, for your continuous questioning of *why* I am doing this, for your amazing supportive *honesty* and *clarity*. I'm not sure how I would have kept walking this path without you on my side. Thank you. To my longest serving educational guide, teacher, mentor, hero, and friend, Dr. Jacqui Kelleher. Jacqui: when I first approached you five years ago to become my MA advisor at Sacred Heart University, I certainly didn't imagine that we would become such good friends and close colleagues, or that we would still be working together five years later! Your expertise in *creativity*, *teaching* and *learning*, and your *intensity* and *integrity* as a professional and as a human being are an example that I hold close to my heart; your ability to tell me I'm being *ridiculous*, while making me feel like the most important person in the world is truly amazing. Thank you for walking this trail with me. To my newest friend and guide, Dr. Kara Sammet. Kara: I deeply appreciate your professional guidance and expertise in instrument development and Rasch modeling, but, above

all else, I value your ability to *push* me as a researcher and human being, to continuously *make me think* about widening and narrowing my perspective. You made me proud to call myself a Raschie! You often don't realize it, but you are the consummate teacher. I hope you persevere in teaching, you truly have a gift. Thank you. Finally, to my expert reader, occasional house guest, mentor, and advisor, Dr. Denise Mitten. Denise: I was so impressed with your reputation before we met; your presence on the faculty was one of the most important factors in my decision to join the program at Prescott College. As we started to work together on that first research methods course, before I knew what my mission would be, I realized that your reputation was merely a reflection of a deep expertise as a teacher-researcher, grounded in real experience of discovering and creating knowledge. When you agreed to become my expert reader, it felt like everything came together for me, closing the circle of my journey into a more complete whole. Thank you.

To the many amazing teachers and school support professionals in the Bridgeport, CT public schools: thank you. You go to work every day in difficult circumstances, and do your human best to make the lives of your students better. To Ms. Mayra Perez, PBIS Team Leader in Bridgeport: Mayra, you have been my research partner, counsellor, and friend on our journey together, this work product is a testament to your expertise and your dedication to your vocation. Thank you for all you do every day. To all the leaders in the Bridgeport schools, particularly to Ms. Susan Smith: thank you for helping me develop the idea of the Climate4Creativity project and bring it to fruition in the Bridgeport schools. Your leadership in challenging times was an inspiration to your many followers and collaborators over the years: I wish you happiness and peace in your well-deserved retirement.

To my church family at Center Congregational Church, United Church of Christ in Manchester, CT. Particularly to director of music, Ms. Andrea Simmons, and the wonderful, dedicated, creative people in Chancel Choir: singing with you every week has been like creative therapy. To Pastor Joyce O. Crutchfield, for asking me how it was going every Sunday, for finding me a secluded place to write in the church library, and just smiling to let me know you care. To my colleagues and students at Albertus Magnus College, for your support and feedback over the last three years. To my research assistant and friend, Ms. Gina DeMartino, for listening to me, telling me to shut up and get on with it, and pushing me forward every day, even when you weren't my research assistant any more.

And, finally, to my family. To my mum, Jane Gardener, for helping me to see how to be true to myself; to my wife, Kerry, for helping with my statistics, and for carrying the parenting load, while I was off writing, and working in Bridgeport and New Haven. And, above everyone else, to the most important human in my life: my daughter, Marion. Marion, my darling, I started this work because of you. Your hugs and smiles, your questions about "my test," your ability to light up the room with your laughter, your elaborate school scenarios, bullying role plays, the drama of kindergarten, first, second, and third grade, all played out with Barbies and the most hardworking school principal in the world: *Aladdin!* Thank you for putting up with grumpy daddy these past couple of years. I love you beyond what humans can measure.

Imagine a ridiculous, overeducated British guy surrounded by a circle of amazing, strong, smart, inspirational women: Truly, I am the luckiest man alive today. Thank you.

Oppression in Language

As a man, I am aware of the oppression implied by the use of certain language in my writing and in my spoken and unspoken word. In this work, it is my intent to use language which implies no oppression, supports the open expression of thoughts and ideas from any source, and embraces all forms of diversity and difference in the lives of human and other than human creatures. To this end, throughout this dissertation, in my own creative work, I use gender-neutral words, such as "humankind" in preference to "mankind," and, "germinal" in preference to "seminal." Where such gender-neutral choices do not exist, I use feminine language in preference to masculine language. For example, I use the feminine pronoun "she," in preference to the masculine pronoun "he." In addition, it is my intent to use "people first" language in all descriptions of people. For example, instead of referring to an "obese child," I would refer to a "child with obesity." This is a purposeful expression of my belief that personhood is core and characteristics of difference are secondary when describing a person.

While it is my intent to take great care in addressing this issue in my own writing, I am aware that many of the writers and researchers quoted in this text were writing at a time when there was less acute awareness of this issue. To quote Michael Crotty, from the preface to his text on the foundations of social science, "Since my readers need no help to recognize and deplore these usages, I have refrained from interrupting the text with *[sic]* many times over to point them out." (Crotty, 1998, p. vii). If there are places in this text where I have failed to achieve my intended standard, I apologize and seek the forbearance of the reader.

CHAPTER ONE:

STARTING AT THE BEGINNING

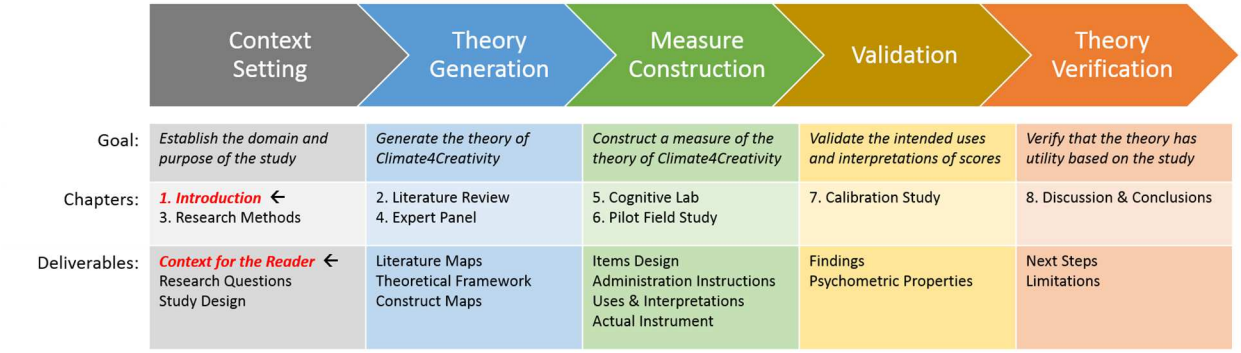


Figure 1. Dissertation Roadmap – Introduction.

This work began as a brainstorming discussion with Susan Smith, the (now retired) head of social work in the Bridgeport, CT public schools. With a referral from my doctoral mentor, Jacqui Kelleher, I had approached Susan to explore what opportunities there might be to work with the Bridgeport schools on a project for my dissertation research. I was determined to take on a project that had value for real world practice in a real school system. At the end of that first meeting in March 2012, Susan introduced me to Mayra Perez, team leader for the Positive Behavior Interventions and Supports (PBIS) team in Bridgeport, and between the three of us, we conceived of an idea that would help solve a problem for the Bridgeport schools, and simultaneously allow me to break some new ground in exploring a topic of deep interest to the field of sustainability education.

The Bridgeport schools had recently lost funding for its annual school climate assessments, conducted by the Yale Child Study Center using the Comer School Climate Instruments (Yale School of Medicine, 2009), and they were not sure how to fill that gap. When

I returned to meet with Susan and Mayra a few weeks later, I brought a proposal to develop and complete a school climate assessment, as mandated by Connecticut antibullying law (An Act Concerning the Strengthening of School Bullying Laws of 2011) combined with an actionable assessment of the learning environment in the schools, with particular emphasis on the environment for teaching and learning creativity. I called the development and implementation of this instrument the Climate4Creativity project. A few weeks later, following a public presentation, the proposal was approved by the Bridgeport Board of Education and the Office of the Superintendent. Work in the Bridgeport schools began in earnest during the dog days of the summer of 2012, and my collaboration with the many excellent professionals in the Bridgeport schools continued through early 2014. This dissertation is a direct result of that collaboration, particularly of my close partnership with Mayra Perez and her team of part time PBIS specialists, social workers, school psychologists, guidance counselors, teachers, and school leaders embedded in the Bridgeport school system.

The Bridgeport Public Schools

Bridgeport, CT public schools were an ideal setting for this work. The Bridgeport schools consist of some 30 unified elementary schools (grades PK-8), along with three large high schools (grades 9-12). The district also operates a number of magnet schools including two inter-district magnet schools accepting students from inside and outside of the district. The Bridgeport schools, located in Connecticut's largest and poorest city, is a large unified urban school district serving the needs of over twenty thousand students.

Historically, Connecticut public schools had been amongst the very best public schools in the U.S., consistently demonstrating superior outcomes with graduating students among the best prepared in the country for career and college success. In many ways because of the state's

position at or near the top of the country's league table of public school systems, adapting to change in practice had been more problematic in Connecticut than in other systems. Educators and education leaders had sometimes responded defensively to change, secure in their belief that Connecticut schools were already wonderful. In the National Assessment of Educational Progress study (Hemphill & Vanneman, 2011), while Connecticut's average standardized test scores continued to be among the best in the nation, Connecticut was ranked toward the bottom in the nation in terms of the educational achievement gap. African American and Hispanic/Latino students demonstrated significantly lower average standardized test scores than their Caucasian counterparts. While achievement gaps of this type were evident in many public school systems, Connecticut's achievement gaps were among the worst. The publication of this study sent shockwaves through the Connecticut system, forcing education leaders to re-evaluate everything from teacher certification to required curriculum and assessment processes to high school graduation requirements.

In the general context of Connecticut public schools, Bridgeport schools were among the most challenged, and, along with Connecticut's two other major cities (Hartford and New Haven) were a major factor in creating the achievement gaps in the state. More than any other district, Bridgeport had a reason and a mandate to improve its schools. And so, in the period from 2012-2014, along with all Connecticut Public Schools, Bridgeport schools were being transformed by state mandate. Unlike many other districts in Connecticut, however, Bridgeport lacked the resources – human, physical, and financial – to implement many of these *unfunded*, *under-resourced* transformations with fidelity. Lack of funding was not an excuse in Bridgeport, but an *ongoing reality* that informed and influenced every management decision in the district. It is in this context that this dissertation study began.

Circles of Sustainability

The field of sustainability is broader than environmentalism, encompassing social as well as physical dimensions of planetary existence. The *Circles of Sustainability* model, for example, discusses four equally important domains of sustainability: economics, ecology, culture, and politics (James, 2015). These domains describe an integrated and balanced framework, designed to measure sustainability of a region, city, or domain. Figure 2 illustrates the circles of sustainability model using a 2012 sustainability assessment conducted in the city of Melbourne, Australia.

This dissertation examines the *elementary schools* within a city, as places where children learn to engage with the four domains of sustainability. Education and learning are located in a number of places within the *culture, politics,* and *economics* domains of the circles of sustainability. If sustainable development requires improvement on a wide variety of integrated perspectives of sustainability (James, 2015), then creation of a *sustainable education* platform is an important consideration.

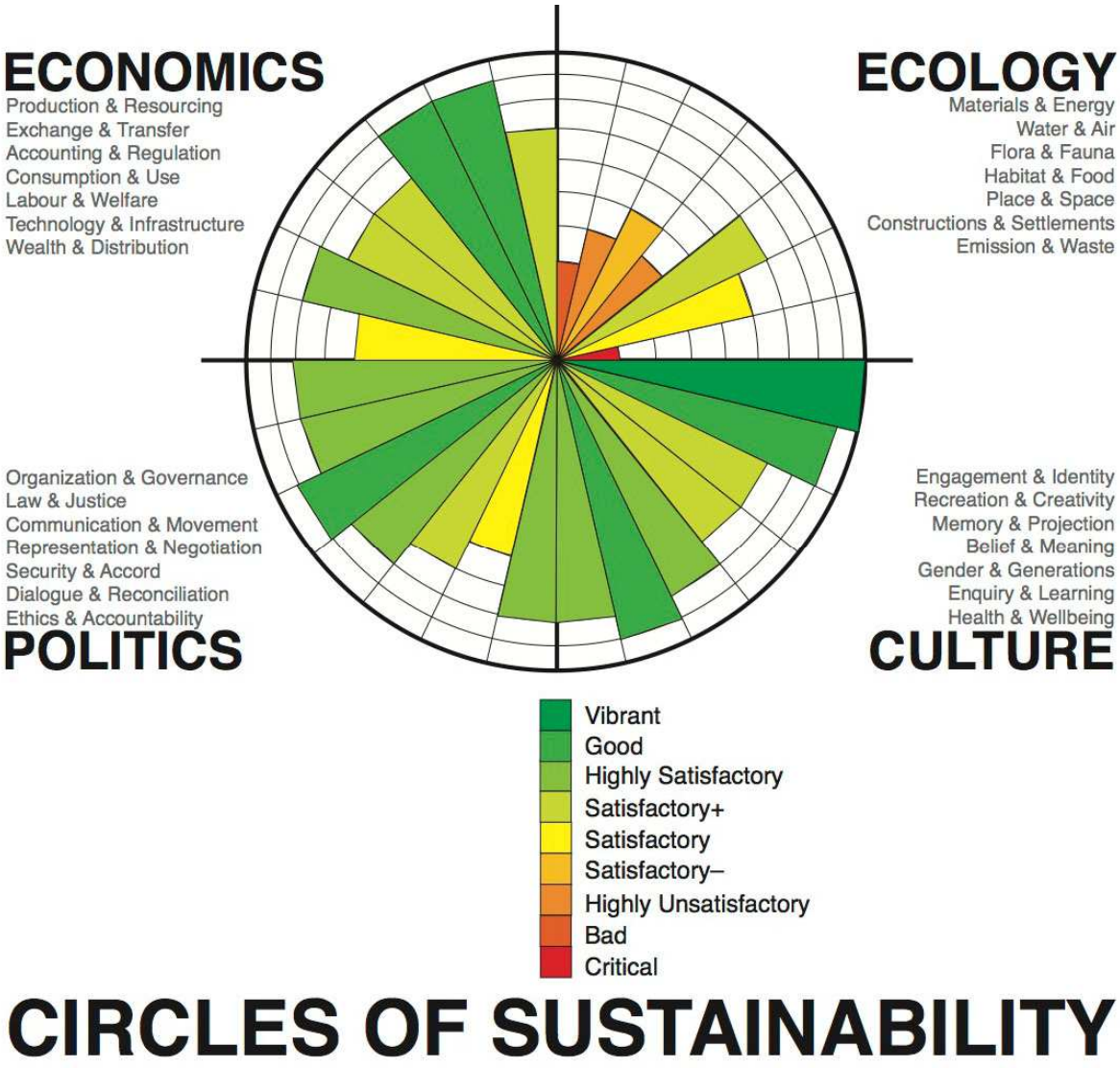


Figure 2. Example of Circles of Sustainability Concept – Melbourne, Australia.

Reproduced under Creative Commons ShareAlike 3.0 License.

Toward a Sustainable Education System

Since John Dewey wrote *Experience and Education* (1938), educational practice and policy have been evaluated on a continuum of transmissive to transformative education. According to Sterling (2001), transmissive education can be characterized as an instructive or training praxis, with an expert teacher communicating (transmitting) a set of pre-defined content

knowledge, using tools such as behavioral change to deliver a rigidly defined list of factual knowledge and skills imposed by an educational hierarchy. Transformative education, in contrast, can be characterized as an iterative process of constructing (transforming) meaning based on authentic experiences. Transformative teachers focus on processes of mutual transformation, delivered in a responsive and dynamic way to deliver a conceptual understanding and increased capacity for learning. Transformative education is based on democratic notions of an ongoing process of appreciation, cooperation, and open-ended enquiry (Sterling, 2001). The continuum of sustainable education is illustrated in Figure 3.

Figure 3. The Sustainable Education Continuum.

By almost any existent measure of success, U.S. public education is not living up to Sterling's (2001) vision of sustainable education. In recent years, the standards movement has succeeded in moving public schools away from improving outcomes for students toward a focus on high-stakes testing of students, using standardized test scores as the primary and often only measure of school quality, teacher quality, and individual academic performance (Ravitsh, 2010). This combined alignment of all school improvement around a single measure of success has caused a significant shift toward transmissive education in U.S. public schools. To sustain

and improve public education over time, a more holistic approach is needed, moving schools back toward a balance of transmissive and transformative practices. This work supports a sustainable, progressive, transformational education system, deliberately designed to build a sustainable culture, and ultimately to support a sustainable global human existence.

Teaching and Learning Creativity

Creativity should not be confused with *artistic work*, although it is clear that many forms of art, music, and poetry require creativity at their highest levels of mastery. The conception of creativity in this work is *the ability to produce work in any domain that is novel, exhibits high quality, and is appropriate to its domain* (Kaufman & Sternberg, 2010). Creativity is inherently challenging: The cognitive act of *creating* is the most complex human cognitive ability in the taxonomy of learning objectives (Anderson & Krathwohl, 2001), placing creativity at the pinnacle of human capability and achievement. The study of human creativity encompasses not only *individual creative ability* but also the *creative process*, the study of *creative products*, and the study of the *creative environment* – the conditions required to enable creativity (Csikszentmihalyi, 1996).

Creativity is not an ability learned overnight. According to Csikszentmihalyi (1996), creativity is the cultural equivalent to biological evolution. In U.S. culture, new ideas and inventions, unlike DNA, are not automatically transmitted to the next generation: Children need to be explicitly *taught* about the creative inventions of their ancestors and *taught* about the creative process. As they *learn* about these ideas, they are *enabled* to create their own ideas, building on and sometimes replacing those that came before. For a person to become creative, she must practice creativity in a simultaneously supportive, challenging, rigorous, and nurturing environment, she must be encouraged to take risks and make mistakes, and in the process, to

purposefully develop her own creative abilities. This special learning environment is needed to deepen creativity in every domain – from art to music to business to mathematics to quantum physics to software engineering to poetry to genetics. Defining, measuring, and ultimately helping school leaders purposefully construct an environment for teaching and learning creativity in an elementary school setting is the strategic goal of this work.

Healthy School Climate

In order to learn, humans need to be able to concentrate. School climate is the sum of the culture, practices, values, people, place, and decisions that define the lived environment of a school. Thapa, Cohen, Guffey, and Higgins-D'Alessandro (2013) defined healthy school climate in terms of five dimensions: safety, relationships, teaching and learning, institutional environment, and the school improvement processes. It is evident that deficits in one or more of these five dimensions would degrade human learning in a school, simply because students struggle to concentrate in a negative school climate. If students fear for their physical safety, if students have negative relationships with their peers or their teachers, if teachers are ineffective, if the physical environment is dangerous or depressing, and if the school is decaying rather than moving forward; if any of these situations exist, student learning will be negatively impacted.

Modern school climate practices, particularly Positive Behavior Interventions and Supports (PBIS) and Social Emotional Learning (SEL) systems, provide evidence-based tools and methods for schools, designed to systematically help school professionals improve school climate over time by focusing on positive, prosocial behavior. School climate is important, because it is the primary approach used in U.S. public schools to manage the school environment to ensure that students can concentrate at school, enabling them to learn.

Significance of this Work

Human understanding of the universe is rapidly expanding. New scientific discoveries, new technologies, new ways of being, and new ways of knowing emerge daily. According to Kurzweil (2005), the rate of paradigm shifting innovation approximately doubles every decade, and the amount of recorded human knowledge approximately doubles every year. This unprecedented explosion of new ideas and new knowledge has created an environment of rapid change and exponential growth in complexity. By the time my third grade daughter graduates high school, a decade from now, the amount of human knowledge will have doubled ten times, making the sheer volume of knowledge 2^{10} or about 1,000 times greater than it is today. This is a mind boggling change in the *quantity* of knowledge in the world. In ten years, the human race will know 1,000 times more than today.

At the same time, the past two decades of U.S. history have seen public schools move away from creative work toward a focus on linearity and rational thought. The impact of this change is evidenced by studies of creative thinking abilities among public school students across the country (Kim, 2011). Along with rational and evaluative processes, creative and reflective practices are essential components of all problem solving, but are particularly important when addressing complex challenges, such as those presented by the expanding universe of human knowledge and the simultaneous challenges created by global warming and human exploitation of limited environmental resources. The rational and convergent thinking skills which are the focus in U.S. public schools today are necessary, but not sufficient, to address the challenges and opportunities of sustainability in 2015, let alone to address the unknown and unforeseen opportunities and challenges of sustainability in the future. It is important to equip children with deeper, more complete thinking strategies that incorporate reflection and creativity into a cohesive, integral, open, dynamic thinking system. This system has to resist teaching children

solely *what* to think, and balance this with a focus on *how* to think more completely. Only by development of creative and reflective practices in our schools can we hope to prepare the next generation to survive and thrive in the uncertain new world we have created for them. This is an important and urgent call for action for school superintendents, principals, and teachers across the U.S. public school system.

The Theory of Action

This call to action will have little utility, unless supported by a rigorous *theory of action* to enable school leaders to make actual progress toward the goal of bringing creative and reflective practices into their schools. Without a theory of action, this work might provide a useful measure of the current state of an elementary school, but cannot achieve any systematic, substantive change in public education. To achieve anything important, to make significant change to any system, particularly in complex and ambiguous situations, the change agent must gain a clear understanding of the system being changed, and apply a singular focus on using tactics to elevate the constraints in the system.

For the purposes of this work, a system is defined as an entity or organism designed to achieve a particular result. Under this definition, it is clear that a U.S. public school is a system designed to achieve learning among its students. Systems are dynamic, can be chaotic, and inevitably include both human and nonhuman, organic and inorganic components. According to Goldratt's theory of constraints (Goldratt, 2010), *any system is restricted by a small number of constraints*. This requires the change agent to measure the *resistance* at each point of constraint within the system then focus all her resources to *elevate* throughput at the most resistant constraint, through carefully designed interventions. Once the most resistant constraint has been elevated, the change agent measures resistance throughout the system and begins the process

over again, in a systematic *process of ongoing improvement* to the system. According to Goldratt (2010), this is the only method to systematically increase the productivity of a complex system.

Inevitably, humans want change to happen quickly. But, meaningful change must happen over a period of time, it cannot happen overnight. This is particularly true when cognitive or mindset shifts are needed among the humans engaged in the system. Once leaders identify new system goals, they must execute a systematic process of ongoing improvement, focused on achieving these goals. This metaframework for systems change does not imply any particular solutions, but requires the humans in the system to situationally apply the best solutions they can develop at a particular point in time. The definition of this change management system *is* the theory of action generated by this research. Figure 4 illustrates this metaframework, using an established model from operations research (Goldratt, 2010) – the measure-focus-act cycle.

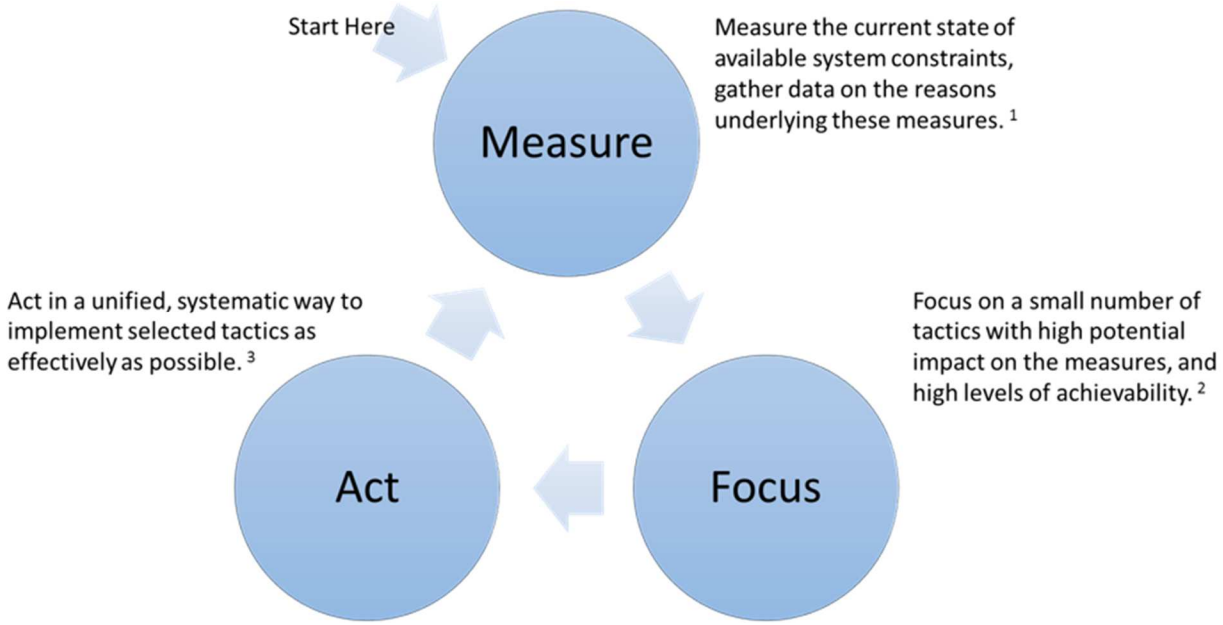


Figure 4. The Climate4Creativity Measure-Focus-Act cycle.

Notes. [1]Measures will change over time, must be repeated periodically, and will likely call for different tactics at each measurement cycle. [2]It will always be more effective to focus on fewer tactics (ideally a single tactic) at any point in time. [3]One action cycle will never be enough to optimize a system, repeated systematic action is required to improve a system over the long term.

Opportunity Statement

By enhancing the practice of school improvement with a specific focus on *teaching and learning creativity*, an opportunity emerges to better prepare young people to participate in the social life of the planet, thereby empowering the sustainability of human society. To achieve this, an action-oriented *theory of climate for creativity* is proposed, a measure of the theory is constructed, and the intended uses and interpretations of its scores are validated. Figure 5 illustrates the opportunity cycle for this work.

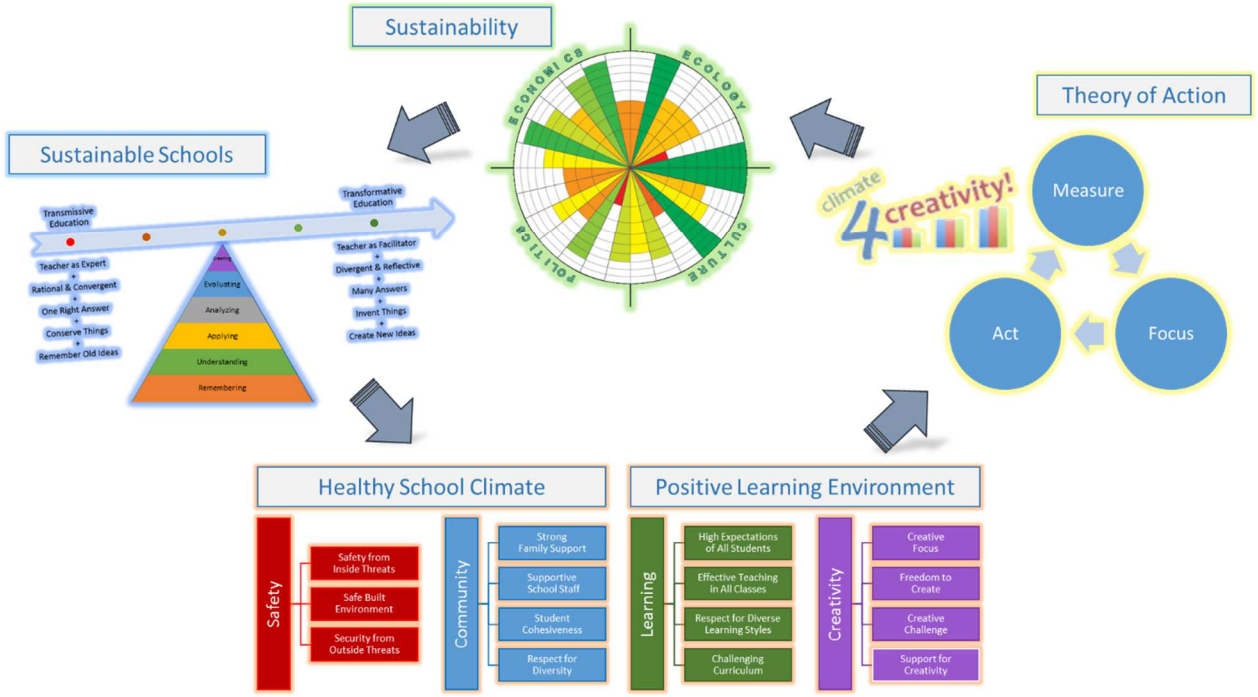


Figure 5. The Climate4Creativity Opportunity Cycle.

Structure of this Dissertation

This dissertation is designed to tell the story of the three year journey into this fascinating domain of research. This first chapter sets context for what follows. *Chapter 2 – Finding Opportunity in the Literature* documents the findings of a systematic literature review of the relevant literature, setting the scene for generation of the theory of Climate4Creativity. *Chapter 3 – Mapping the Methodological Road Ahead* documents the research questions, and corresponding research design for the Climate4Creativity project in the Bridgeport, CT public schools. *Chapter 4 – Gathering Expert Perspectives* describes the first study in the research design, and documents the Theory of Climate4Creativity through a set of construct maps based on the findings from the study. *Chapter 5 – Understanding the Target Respondent* describes the second study in the research design, a cognitive lab, used to develop and refine the items to be

used in the Climate4Creativity elementary and middle school instruments. *Chapter 6 – Piloting the Instrument*, describes the pilot field study conducted in one middle school in Bridgeport, CT, and used to refine the administration instructions and better understand the intended uses and interpretations of the scores from the instrument. *Chapter 7 – Calibrating the Measure* describes the calibration study, conducted in twenty five elementary schools in Bridgeport, CT, and provides thoughtful evidence of validation of the intended uses and interpretations of the scores generated by the measure. *Chapter 8 – Discussion and Conclusions*, discusses the successes and limitations of the Climate4Creativity project, discusses how this type of instrument may be used to drive real change in a school setting, connects the findings from the research back to the context for the study, and identifies next steps to move this work forward. The figures at the start of each chapter locate the content in that chapter within this journey.

CHAPTER TWO:

FINDING OPPORTUNITY IN THE LITERATURE

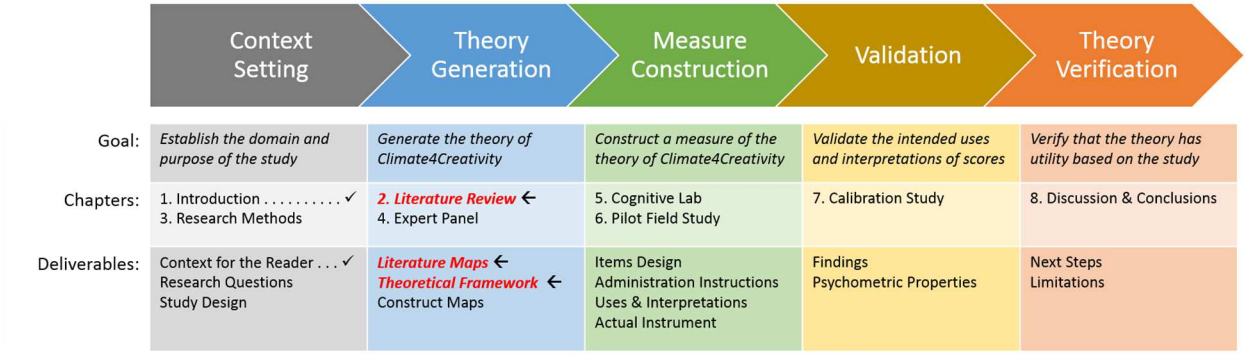


Figure 6. Dissertation Roadmap – Literature Review.

By enhancing the practice of school improvement with a specific focus on *teaching and learning creativity*, an opportunity emerges to better prepare young people to participate in the social life of the planet, thereby empowering the sustainability of human society. To achieve this, an action-oriented *theory of climate for creativity* is proposed, a measure of the theory is constructed, and the intended uses and interpretations of its scores are validated.

Finding Opportunity in the Literature

The literature relevant to elementary school (grades 3-8) student perspectives on healthy school climate, positive learning environment and the teaching and learning of creativity is broad and diverse. This chapter provides a summary of those topics most relevant to the problem being studied in this program of research. Specifically this chapter addresses, 1) The underlying purposes of public schooling, from its inception in the common schools of Massachusetts in the 1800s, through modern school reform; and, 2) The structures and processes needed to create a

schoolwide climate and classroom learning environment conducive to the teaching and learning of creativity in a public elementary school setting. Definitions of the goals of education, creativity and higher order thinking, learning environments, and school climate are provided, and the empirical and theoretical bases for these definitions are reviewed.

The search-appraisal-synthesis-analysis (SALSA) systematic literature review methodology (Booth, Papaopannou, & Sutton, 2012) was used to search for candidate articles and texts, appraise each identified article or text, synthesize the articles into a coherent structure, and analyze the relationships between the components of the literature. This methodology begins by selecting credible sources, and defining search terms within those sources, then uses a structured rubric-based criterion for scoring each article or text, then synthesizes and analyzes each article or text, using graphical tools to understand the structure of the literature, in addition to the content of each relevant article or text. The written output is designed to carefully document the most relevant research, but the methodology is designed to provide a broad analysis of the literature, even if specific articles and texts are not included in the final write-up. A complete description of the methodology used in this literature review can be found in Appendix B.

Context: The Promise and Problem of Public Schooling

Jefferson's Vision. The primary purposes of public education were clear as early as the writings of Thomas Jefferson, when he wrote to his friend Charles Yancey, in 1816,

> if a nation expects to be ignorant & free, in a state of civilisation, it expects what never was & never will be. the functionaries of every government have propensities to command at will the liberty & property of their constituents. there is no safe deposit for these but with the people themselves; nor can they be safe with them without

> information. where the press is free and every man able to read, all is safe. (Jefferson, 1816)

With these profound words, Jefferson crafted the argument on which the principle of free public education is based – an idea he advocated throughout his career in public service. When Horace Mann became the nation's first Secretary of Education, appointed in the State of Massachusetts thirty years later, he created a clear vision for what were then called *common schools*. These schools would systematically serve all boys and girls with a common body of knowledge: to Mann, public education was a critical part of the sustainability of society and an essential part of the systematic operation of a democratic nation,

> Education, then, beyond all other devices of human origin, is the great equalizer of the conditions of men--the balance-wheel of the social machinery… And hence it is, that the establishment of a republican government, without well-appointed and efficient means for the universal education of the people, is the most rash and fool-hardy experiment ever tried by man. (Mann, 1848)

Over time, the Massachusetts common school movement led directly to the creation of state public school systems across the United States. Despite the forward momentum provided by these early activists, it was not until the end of the civil war that all states were *required* to include free public education *in their constitutions* as one of the conditions for re-integration into the United States. By 1900 almost all children attended public elementary schools, and by mid-century nearly 80% of teenagers attended public high schools (Mondale & Patton, 2001).

Bloom's Taxonomy of Learning. Over the years since the creation of U.S. public schools, much research has been conducted, in support of and guided by these dual principles of developing informed citizens and educating the workforce. In 1956, one of the most important

research based texts on the objectives of education was published, edited by Benjamin Bloom. This germinal text (Bloom, 1956), examined the cognitive domain of learning, providing a new, clearer way of thinking about teaching and learning, and creating what would become the definitive taxonomy of the goals of education (for a detailed discussion of Bloom's taxonomy in relation to other frameworks of educational goals, see chapter 15 of Anderson & Krathwohl, 2001). *Bloom's Taxonomy*, as it came to be known, has been embedded in educational practice and regulation ever since, emerging as one of the most influential educational texts ever written (for example see Anderson & Krathwohl, 2001; Shane, 1981). The six levels in the original taxonomy – knowledge, comprehension, application, analysis, synthesis, and evaluation – continue to be used as the basis for teacher preparation, curriculum development, and testing today, despite criticism of some aspects of the taxonomy.

Writing in the original (1956) text, taxonomy team members expressed concerns that the taxonomy might, itself, become a trap – causing teachers to use the taxonomy to check boxes, rather than to deeply understand their practices and the learning experiences they were providing to their students (Bloom, 1956, pp. 5-6). In addition, like many researchers at the time, Bloom and team made some uncritical assumptions in their work, several of which turned out to be unwarranted. For example, Bloom assumed that the levels in the taxonomy were cumulative and mutually exclusive – that is, in order to master a higher level category, students must first master all lower level categories (Bloom, 1956, pp. 18-19). These issues led to a perpetuation of the belief among teachers that students must remember all the facts of a domain first, before they begin to comprehend and then apply their knowledge in the domain. While subsequent research provided only limited empirical evidence in support of the cumulative aspect for the three middle categories (comprehension, application, and analysis), the entire definition of the knowledge

category, along with the sequencing of the synthesis and evaluation domains proved to not be well supported by the empirical evidence. Research later established that the knowledge domain is significantly more complicated than the definition provided in the original 1956 taxonomy (for a full discussion of these issues, see Kreitzer & Madaus, 1994).

New Bloom. Partly in response to these issues, as well as in response to the evolving context of how children learn and how teachers teach, a new team was formed on the 40th anniversary of the original publication of the taxonomy. The new team included David Krathwohl, a member of Bloom's original team, and was led by Lorin Anderson, a scholar of teaching and learning. The goal of this new team was to revise, update, and improve Bloom's taxonomy based on new understandings of educational and cognitive processes (Anderson & Krathwohl, 2001). The new team clarified and restated the taxonomy, using clearer and more contemporary language, and made some important shifts in the structure of the taxonomy based on new research. The new revision is known as *New Bloom*. The most important shifts from the 1956 taxonomy to the 2001 taxonomy are illustrated in Figure 7.

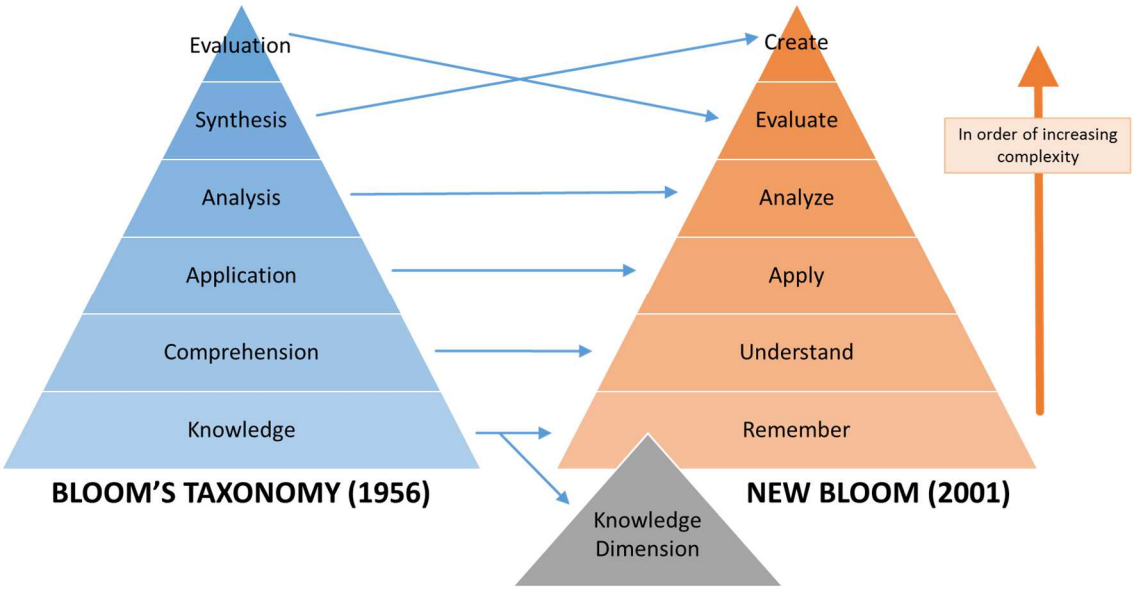


Figure 7. The Shift in Bloom's Taxonomy of Educational Objectives.
Adapted from Anderson & Krathwohl (2001, p. 310).

Most notably, in the new framework, *creating* is moved from a subcategory of Bloom's original *synthesis* level to a new position at the top of the hierarchy. This move, reflecting new research, positions creativity as the most complex and challenging cognitive process embodied by the educational system. This change aligned the new taxonomy with other research into cognition, and clearly supported the societal shift toward the new 21st century era of innovation and change, away from the 19th century era of mass production and stability (Trilling & Fadel, 2009). In addition, the original knowledge category was simplified into a *remembering* category, and an entirely new domain of the *structure of knowledge* was created. In this new knowledge domain, a hierarchy of knowledge exists as a parallel dimension alongside the cognitive process dimension, implying different categories of knowledge as well as different cognitive processes.

Important outcomes of education, such as critical thinking, problem solving, and creativity, are not specifically called out in either the 1956 or 2001 taxonomy. This is because

each of these must, of necessity, span multiple levels of the hierarchy. For example, problem solving requires experience and skill in several of the levels of the taxonomy – a problem solver must be able to remember important facts, understand the implications of these facts, apply processes to analyze and evaluate potential solution paths, and ultimately must have the ability to both evaluate existing solution options and create new solutions (Anderson & Krathwohl, 2001, pp. 311-312). New Bloom forms the underlying framework of learning objectives for this dissertation.

The Problem of Contemporary School Reform. According to many popular authors, and indeed much contemporary educational research, U.S. public schools are close to breaking point, suffering from a dual challenge of increasing demands, and decreasing availability of resources to accomplish those demands. Ravitsh (2010) argued that the modern standards movement had moved from a focus on improving outcomes for students to a focus on high-stakes testing alone, using standardized test scores as the primary and often only measure of school quality, teacher quality, and individual academic performance. She went on to describe how the accountability and parental choice movements, initially focused on improving school performance for all schools, subsequently aligned themselves behind these same measures. This alignment further amplified the problems of challenged school systems. Those families with sufficient resources tended to move their children from challenged schools to choice schools, leading to an inevitable cycle of decline in the challenged schools and improvement in the choice schools, which continued to enjoy higher levels of parental support, funding, and ability to focus on their students (Ravitsh, 2010).

This intense national focus on *common core standards,* and assessments aligned with common core standards, were more problematic in Connecticut than in many states.

Connecticut's constitution specifically extended the U.S. Constitution by placing a primary duty on the public education system to prepare citizens for participation in democratic processes, with a secondary duty to prepare graduates for work and higher education ("Connecticut Coalition for Justice in Education Funding, Inc., v. Governor M. Jodi Rell et al.," 2010). Common core, on the other hand, had a singular purpose to prepare students for higher education and work (Common Core Standards Initiative, 2012). The broader constitutional duty of Connecticut public schools presented educators with a dilemma: Common core and associated assessments were designed to measure only what was required for mastery of work skills and skills needed for entry into higher education.

Much has been written on the impact of the standards movement on educational outcomes in the U.S. From criticism of a lack of focus on excellence (Gentry, 2006), to articles in the popular press analyzing the boom in private tutoring and test prep (Sullivan, 2011), to research into the impact of the shift away from creativity toward test preparation (Kim, 2011). In this last research into creative thinking abilities among school children in U.S. public schools, Kim (2011) analyzed the historical records from normative tests of the figural components of the Torrance Tests of Creative Thinking (TTCT), and concluded that, based on these scores, on almost every figural measure in the TTCT,

> over the last 30 years, 1) people of all ages, kindergartners through adults, have been steadily losing their ability to elaborate upon ideas and detailed and reflective thinking *[sic]*; 2) people are less motivated to be creative; and 3) creativity is less encouraged by home, school, and society overall. (Kim, 2011, p. 292)

This research into creativity in schools, particularly at the elementary school level, provides an important context for this dissertation: In a society shifting toward more and more innovation

and creativity (Trilling & Fadel, 2009), in an educational environment where the research calls for more focus on creativity as a goal of education (Anderson & Krathwohl, 2001), it seems that the last thirty years of school change have created a generation of graduates *less* skilled and *less* experienced in creative practices, rather than one more skilled and more experienced in creative practices (Kim, 2011).

Thematic Structure of the Literature

The literature on school climate and the environment for teaching and learning creativity fell into four major themes, with fifteen subthemes, as described in Table 1.

Table 1.
Major Themes Identified in the Literature on School Climate, Learning Environments, and Creativity

Theme	Description	Subthemes
School Safety	Research around physical and emotional safety at school, including bullying and anti-bullying practice, physical facilities, and site security.	Safety from Inside Threats
		Safe Built Environment
		Security from Outside Threats
School Community	Research around the connectedness and cohesion of the school community, including family and school staff engagement, student connectedness, and respect for diversity.	Family Support
		Supportive School Staff
		Student Cohesiveness
		Respect for Diversity
Learning Environment	Research around the learning environment within a school, including expectations of learners and teachers, diversity of content and activities, effective classroom management, and work difficulty.	High Expectations
		Effective Teaching
		Diversity of Learning Styles
		Challenging Curriculum
Creativity & Higher Order Thinking	Research around the enablers and barriers to learning creativity and higher order thinking.	Creative Focus
		Freedom to Create
		Creative Challenge
		Support for Creativity

The themes and subthemes within the literature were assessed for *depth*, and the connections between findings were assessed for *strength*. These theme depths and connection

strengths were translated into graphical form in Figure 8. Larger bubbles indicate more depth to the literature, both in terms of the quantity of articles addressing the themes, and my assessment of the quality of insight within these articles. My qualitative assessment of the strength of relationships between themes is indicated by the weight of the lines connecting themes and subthemes, and is reflective of the degree to which the authors of articles and texts in the review connected the themes in their work.

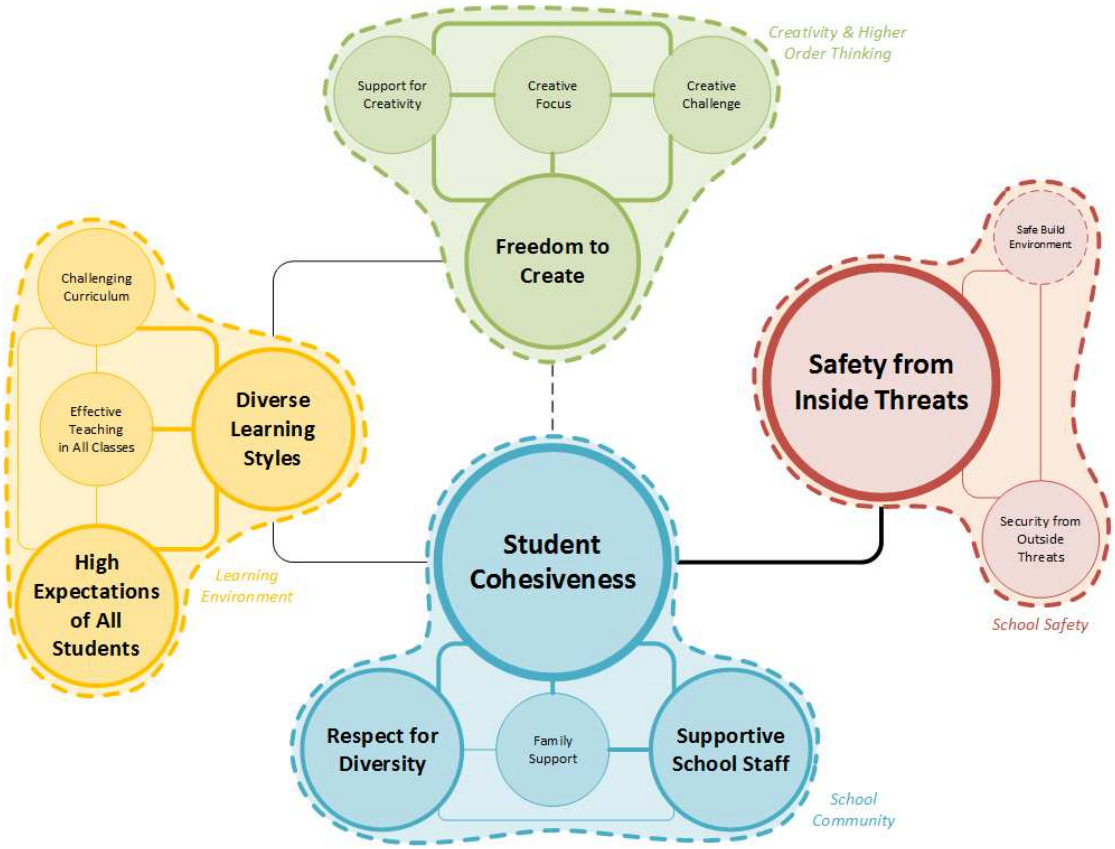


Figure 8. Thematic Structure Map of the Literature on School Climate, Learning Environments, and Creativity

There was evidence in the literature of close connections between school safety and school community. Clearly, if students feel connected to their school, families and staff are supportive of the school, and diversity is respected, safety from inside threats is directly affected.

The security from outside threats and safe built environment subthemes are also somewhat connected to all four of the subthemes within the school community theme. This implies that school community has a kind of protective effect on school safety overall, and safer schools in turn provide a space for school community to flourish. There was evidence of relationships between school community and learning environments; and between learning environments and creativity, but little evidence of relationships between school community and creativity. No evidence was found in the literature of direct relationships between school safety and creativity.

Underlying Theories in the Literature

A series of important theoretical perspectives were identified from the literature, and are collected in Table 2. Each major theory serves a specific purpose in this study.

Table 2.
Major Theories Identified in the Literature on School Climate, Learning Environments, and Creativity

Theorist(s): Theory	**Components**	**Description**	**Purpose in this Study**
New Bloom: Educational Learning Objectives	Creating Evaluating Analyzing Applying Understanding Remembering	A hierarchy of learning goals, sequenced by increasing complexity.	Establishes creativity as the pinnacle of human learning, and higher order thinking as essential components of any learning system.
N.S.C.C.: School Climate	Safety Community Learning	A system of interventions and measures to promote prosocial behaviors in school contexts.	Provides a model of how learning should be delivered on a schoolwide basis.
Cziksentmihalyi: Flow Theory of Creativity	Domain Field Person	A systems perspective on learning and manifesting creativity based on the domain, field, and individual creator.	Provides a framework for defining the detailed attitudes and characteristics required to achieve creativity in a school.
Goldratt: Theory of Constraints	Measure Focus Act Repeat	An empirical systems approach to strategic, ongoing improvement in a complex systems environment.	Provides a thinking framework for development of a system of strategic school improvement.

Positive School Climate: An Unexpected Antidote to Standardized Testing?

In their recent review of school climate research, Thapa et al. (2013) proposed that researchers and educators adopt a common definition of positive school climate, based on the work of the National School Climate Council, a national advisory body formed to press for school climate improvement (National School Climate Council, 2007), as follows:

> A sustainable, positive school climate fosters youth development and learning necessary for a productive, contributive, and satisfying life in a democratic society. This climate includes norms, values, and expectations that support people feeling socially, emotionally

> and physically safe. People are engaged and respected. Students, families and educators work together to develop, live, and contribute to a shared school vision. Educators model and nurture an attitude that emphasizes the benefits of, and satisfaction from, learning. Each person contributes to the operations of the school as well as the care of the physical environment. (p. 4)

This definition asserts a theory that by creating a particular school culture and organization – founded in shared values, norms, and goals – positive outcomes will be gained in terms of learning, citizenship, school operations, and the physical environment of the school. The roles of educators, students, and families are clearly articulated, and the three primary concerns of school climate; namely *safety*, *community*, and *learning* are laid out. This provides a potential articulation of the role of school climate as a *balancing measure* in public education, providing a framework for development of a democratic society. If standards and standardized testing are designed to prepare students for work and higher education, then positive school climate could be designed to prepare students for participation in a democratic society. If we accept that standards form a clear definition of *what* society expects public school students to learn, then positive school climate provides us with a definition of *how* society expects public schools to deliver this learning to young people in the U.S.

The Impact of Positive School Climate. A variety of research supports the perspective that improved overall school climate has a direct effect on improved student outcomes such as academic results, improved health, reduced dropout rates, lower levels of school violence, etc. For example, Cornell, Gregory, Huang, and Fan (2013) found that classroom prevalence of bullying and aggression in 9th grade was correlated with high school dropout rate, four years later. The implication of this finding is that interventions that reduce bullying and peer

victimization will lead to later improvement in dropout rate. This research is significant because the researchers carefully controlled for a large variety of compounding variables, including school size, socioeconomic status, racial mix, crime rates, and standardized test outcomes. Astor, Guerra, and Van Acker (2010) showed that a lack of feelings of safety in a school is largely associated with breakdowns in relationship among students and between students and adults in a school and with the physical conditions in the school. Koth, Bradshaw, and Leaf (2008) focused on student perceptions of order and discipline and engagement motivation. They found that individual (race, gender, academic history), class level (class size, teacher practices), and school wide factors (school size, teacher turnover) all have influence on overall school climate perceptions of 5th grade students in public school. The implication of this is that both *measures* of school climate and *strategies* designed to improve school climate need to employ multilevel perspectives.

School Safety

Safety from Inside Threats. Most school safety research focuses on precursors, impacts, and prevention tactics for threats to safety driven by the actions of students or other individuals who are inside the school. By examining safety from inside threats from a variety of perspectives, researchers hope to provide educators with both understanding of phenomenon and evidence-based interventions and tactics to reduce the prevalence of bullying, victimization, violence, aggression, and harassment in school settings.

In their three year mixed methods study of nine schools in Israel, Astor, Benbenishty, and Estrada (2009) assessed the characteristics of what the authors called atypical schools, that is, schools which exhibit very different levels of inside threat compared to the communities they serve. The schools studied included a mixture of Arab and Jewish schools that had much higher

levels of violence than surrounding communities, schools that had much lower levels of violence than surrounding communities, and *change* schools where some major shift in levels of violence was evident from their initial quantitative data analysis. Having identified nine schools for deeper analysis, the authors used a variety of collaborative qualitative research techniques (including participant observation, depth interviews, focus groups, hot spot mapping, and photo-elicitation) over three school years to analyze multiple aspects of each school and were able to construct a sophisticated model of what made these schools atypical, essentially,

> Our qualitative findings indicate that the theoretically atypically low violence schools have a combination of organizational and social climate school factors that are implemented effectively by staff and under the leadership of a visionary principal. . . . The overall strength of school social climate and organization emerged as the core variables distinguishing both violent and nonviolent school settings. This included the nature of teacher-child relationships, the presence of clear procedures (carried out by staff who believe in the effectiveness of these procedures), an articulated school safety approach that fit with the academic and social mission of the school, a strong school educational mission, and the use of indigenous cultural values to convey the schools safety mission. When taken together, these effects appeared to be very strong across all school types and cultural settings. (Astor et al., 2009, pp. 440-442)

These mixed methods findings illustrate the importance of holistic approaches to reducing violence, and provide evidence that balanced *organizational* and *social* initiatives, if implemented well, can have a significant effect. It is not sufficient to have written procedures for dealing with violence, school staff have to believe that the procedures will actually work.

An important finding of this research was the emphasis, in the low violence schools, on supportive, nurturing leadership versus disciplinary, enforcement focus. The researchers concluded that harsh enforcement was, in fact, a *contributing cause* of high violence school climates, while caring engagement and strong student-teacher relationships was a *primary cause* of low violence school climates. This was evidenced in their single study findings, and, more significantly, demonstrated in schools where a new principal had implemented either more harsh or more nurturing approaches to school management. In both of these cases, corresponding changes in school climate were observed *following* implementation of this change in leadership style. In this research, changing leadership style was the *only* reason identified for shifting levels of school violence in the change schools.

> The high violence schools had many observable and reported instances of principals and teachers yelling, reprimanding, and unfairly punishing, lecturing, or ignoring the students. The principals in these schools seemed more detached from the student and teacher body. (Astor et al., 2009, p. 444)

Similarly, a 2010 large scale study of authoritative school discipline in the U.S. by Gregory et al. (2010) used quantitative data and analysis to determine that neither solely authoritative regimes nor solely supportive regimes lead to sustainable improvements in school climate. They conclude, instead, that to improve school climate, a deliberate combination of structure and support is required – getting tough and giving support in equal measure – rather than a choice between enforcement and nurturing mindsets. This finding triangulates with Astor et al. (2009), despite different cultural contexts, and diverse research approaches. In their review of school discipline practices, Osher, Bear, Sprague, and Doyle (2010), suggested expansion of the definition of school discipline to encompass ideas of self-discipline and control, emotional

and social learning, and positive behavior. They connected school wide positive behavior supports (SWPBS), social and emotional learning systems (SEL), and effective classroom management practices into a cohesive integrated approach which seems consistent with the research described above. In their comparative experimental study of exclusionary discipline practices and classroom level positive behavior supports (PBS) in an elementary school setting, Mitchell and Bradshaw (2013) showed that PBS supports were associated with improved teacher-student relations, higher actual levels of discipline and order in the classroom, and increased student perceptions of school climate.

School wide practices and expectations are important to provide both an organizational context for school staff and social/behavioral guideposts for students. In their large scale study of aggression in middle school populations, Henry, Farrell, Schoeny, Tolan, and Dymnicki (2011) showed that school norms against violence, expectations of, and actual presence of, positive teacher-student and positive student-student relationships, and school responsiveness to violence and aggression were significantly related to levels of aggression in middle school. Frey, Hirschstein, Edstrom, and Snell (2009) conducted a two year longitudinal, experimental study of elementary students using a schoolwide anti-bullying program ("Steps to Respect") as treatment. The study concluded that implementation of the program led to reductions in violence, aggression, acting out behavior, and even argumentative behavior. Overall the literature concludes that positive schoolwide climate is related to lower levels of aggression among elementary and middle school students.

Recent research has examined classroom level influences on safety from inside threats. In Austria, an analysis by Bergsmann, Van De Schoot, Schober, Finsterwald, and Spiel (2013) showed that classroom structures had a predictive effect on incidents of verbal and physical

aggression. Specifically, this research demonstrated that positive classroom organization led to lower levels of aggression nine months later. In their study of British schools, Pryce and Frederickson (2013) showed that victimization and bullying can be influenced in elementary school students through deliberate construction of a shared commitment to non-violence and effective anti-bullying practices and Veenstra, Lindenberg, Huitsing, Sainio, and Salmivalli (2014) demonstrated that teachers can have a significant effect on incidents of bullying through continuous, persistent, clear articulation of behavioral norms and expectations.

Empirical research has also focused on individual and group precursors to bullying and aggressive behavior in school settings. Bouman et al. (2012) studied the social constructs in Dutch schools which might lead to middle school student aggression and victimization behavior. Their study found that students labeled as persistent bullies by their peers had higher popularity ratings and higher social standing, moderated by lower likeability ratings. Students reported by their peers as persistent bullies did not typically identify themselves as bullies, although they did see themselves as having higher social standing. Students who reported higher levels of victimization showed lower social standing and popularity ratings, and were less liked by their peers. This research implies that bullies and aggressors may be motivated by a desire for social standing and popularity. Similarly, Olthof, Goossens, Vermande, Aleva, and van der Meulen (2011) found that Dutch middle school students used acts of bullying deliberately to increase their social standing. Bullying and violence are not random acts by *bad kids*, from the bullies' perspectives, they are often rational strategies to develop social standing in a school context

Goldweber, Waasdorp, and Bradshaw (2013) used a latent class analysis with large samples of middle school and high school students to study the links between bullying behaviors and feelings of safety and belonging at school, showing a continuum of involvement in bullying

(on a scale from low involvement through verbal involvement to high involvement), and demonstrating that students who exhibit higher bullying behaviors (either as bullies or as bully/victims) tended to have *lower* feelings of safety and belonging at school themselves. This finding implies that students who feel less safe at school may in fact be at risk of *exhibiting* bullying behaviors over time – feelings of safety are not only a symptom of a climate for bullying, but may be a contributing cause to such a climate. Improvements in *perceptions* of safety may, themselves, reduce incidence of bullying. In an earlier latent class analysis, Waasdorp and Bradshaw (2011) examined the impact of frequent victimization in middle and high school students, relating their response choices to later socialization problems, and extreme internalizing and externalizing behaviors. Victims of frequent bullying may find themselves further victimized because of their response to this frequent victimization. This finding calls for early supportive interventions to ensure that such victims do not enter a vicious circle of further victimization. A third latent class analysis, conducted by Wang, Iannotti, and Luk (2012), added rumor spreading, cyberbullying, and other emerging bullying practices to the literature, concluding that physical aggression was more prevalent in 5th and 6th grade, with relational aggression including cyberbullying becoming more prevalent in 7th and 8th grade. They further concluded that most cyberbullying is conducted by a small group of aggressive adolescents who simultaneously conduct both relational and physical attacks on their victims. Cyberbullying is a hyper aggressive form of relational aggression and should be viewed as both an issue in itself and a precursor to physical aggression and violence.

Jutengren, Kerr, and Stattin (2011) examined the effect of peer victimization using structural equation modeling, and showed that Swedish adolescents had increased risk of self-harm, up to and including suicide, when they had been subjected to peer victimization. Strong

positive teacher-child relationships prior to adolescence had a protective effect on both internalizing behaviors and externalizing behaviors, and thereby on incidents of self-harm. O'Connor, Dearing, and Collins (2010) found similar effects in the U.S., concluding that strong positive teacher-child behaviors changed the trajectories of youth development, predicting reduction in externalizing behaviors and protecting students from self-harm.

A recent study by Dukes, Stein, and Zane (2010) examined the relationship between gender, incidents of relational and physical aggression, and weapon carrying in school settings. Their research concluded that girls exhibited higher levels of *relational aggression* than boys, and boys exhibited higher levels of *physical aggression* and higher levels of weapon carrying. An earlier research project, by Estell, Farmer, Pearl, Van Acker, and Rodkin (2008), examined early indicators of aggression in 3rd grade girls, showing that the precursors of aggressive behavior were evident even as early as age 8 or 9. The research team used a cluster analysis to develop a typology of aggressive potential, mapping girls' aggression and popularity. One important finding was that while aggression is a driver of popularity among boys, it does not seem to drive popularity to the same degree among girls. Popularity seemed independent of aggressiveness in girls. In addition, although positive classroom climate did not influence the typology (i.e., the approximate proportion of girls in each category was not significantly different across different classroom climates), classroom climate did influence the *behavior* of the more aggressive girls, moderating their levels of aggression. The implication of these research studies is that gender-focused intervention tactics are possible and are likely required to reduce aggression overall.

Early work by a number of researchers on bullying and school climate examined who are the victims of bullying behavior, demonstrating that a wide variety of *perceived differences*,

including race (for example, Juvonen, Graham, & Schuster, 2003; Nansel, Haynie, & Simons-Morton, 2003), disability (for example, Little, 2002; Marini, Fairbairn, & Zuber, 2001; Norwich & Kelly, 2004; Woods & Wolke, 2004), obesity (for example, Janssen, Craig, Boyce, & Pickett, 2004), and personality traits that don't fit in (for example, Hoover, Oliver, & Thomson, 1993; Olweus, 1995). More recently, lesbian, gay, bisexual, transgender, and questioning (LGBTQ) status has emerged as a targeted difference (for example, Robinson & Espelage, 2011; Scherr, 2012). The consensus among the literature is that *any visible or behavioral difference* might provide opportunities for victimization (Swearer, Espelage, Vaillancourt, & Hymel, 2010).

Other research has sought to understand peer intervention behavior on the premise that peer intervention might be a powerful way to prevent or minimize aggression in school contexts. This peer intervention behavior is becoming known as *upstander* behavior contrasted with *bystander* behavior (Eyman & Cohen, 2009). In their quantitative analysis of middle school age students in the U.S., Batanova, Espelage, and Rao (2014), found that students who had prior experience of being victims of bullying or aggression were more willing to intervene to help other students. Similarly, Nickerson, Mele, and Princiotta (2008), in their small scale analysis of middle school upstanders and bystanders, found that empathy for others accounted for much of the variability between intervention and escape behaviors, regardless of gender. The implication here is that upstander students might be characteristically more empathetic or might become empathetic toward other victims through their own experience of victimization.

Following an horrific school shooting California in 2001, Syvertsen, Flanagan, and Stout (2009) studied middle school students' willingness to either take direct action to intervene, tell an adult, tell a non-adult friend, or do nothing, in a case where they became aware of a dangerous plan by one of their classmates. Knowing that at least twenty classmates knew about Charles

"Andy" Willams' plan to shoot up his school, researchers created a series of scenarios to examine the impact of different school climates on students' willingness to break the code of silence about risky behavior by a classmate (behavioral scenarios ranged from personal drug or alcohol use through school shootings). Overall, students favored taking action on their own rather than using other response strategies, and students with a higher sense of belonging and connectedness to their school were significantly more likely to take action (either directly or by telling an adult) than were those who did not have a high sense of connectedness to their school. This finding reinforced the need for what the authors called *democratic authority* in the school. Pryce and Frederickson (2013) reached an early conclusion, three months into an anti-bullying intervention, that middle school students become more able to control their own bullying behavior when teachers and other adults listen to them and take action on their behalf. Similarly, Eliot, Cornell, Gregory, and Fan (2010) found that 9th graders in Virginia were significantly more likely to seek help from adults in a school, when threatened with aggression and violence, if they perceived their teachers and school leaders as supportive and attentive toward their needs. Students were more likely to report or take action in schools where they felt they had a voice, where they trusted and respected their teachers and other school staff. In contrast, students attending zero-tolerance schools, where enforcement of consequences was swift and automatic, were more likely to feel disconnected from the adults in the school community, and therefore were reluctant to report when threatened with aggression and violence.

Safe Built Environment. In their mixed methods analysis of atypical violence schools in Israel, Astor et al. (2009), concluded that the physical environment of a school grounds and buildings was a significant determinant of degrees of violence and victimization. For example, their photo elicitation research combined with violence hot spot mapping clearly demonstrated

that neglected, invisible, unsupervised places on school grounds were associated with higher levels of violence and victimization, whereas well cared for, clean, supervised, and visible places were associated with lower levels of violence. This has important implications for management of safety on school grounds.

It is clear that non-classroom settings are important contexts for bullying and aggression. A significant body of research over time has demonstrated that up to half of all behavior problems at school occur in non-classroom settings (for example, Astor & Meyer, 2001; Astor, Meyer, & Pitner, 2001; Colvin, Sugai, Good, & Lee, 1997). More recently, McCurdy, Lannie, and Barnabas (2009) found that lunch room supervisors could have a significant impact on cafeteria aggression during lunch period through the use of a constructive, positive game – called the Lunchtime Behavior Game. This finding may provide a model for recess, lunch, and before and after school activities using experiential games. Interesting recommendations about physical spaces were identified in Davies et al. (2013). They recommended creating open spaces for learning, with clean, safe spaces inside classrooms and located in other areas of the school. These collaborative spaces provide opportunities for small group work, outside class work, and could provide safe zones within a school to promote the physical and emotional safety of students. In addition, these spaces were shown to provide a useful resource in development of spaces for creativity and learning.

In their review of U.S. school safety data, Mayer and Furlong (2010) called for a national 10 year strategy to address school safety in comprehensive, evidence based ways. They made it clear that, despite media reports, incidents of outside threats remain very rare, but, their impact can be devastating on individuals and on the communities where they occur. The researchers contrast this with almost continuous low level aggression, negative behavior and bullying, which

has less visible, but in their view, no less important negative impact on individuals and communities. Figure 9 illustrates the structure and context of school safety, according to Mayer and Furlong (2010).

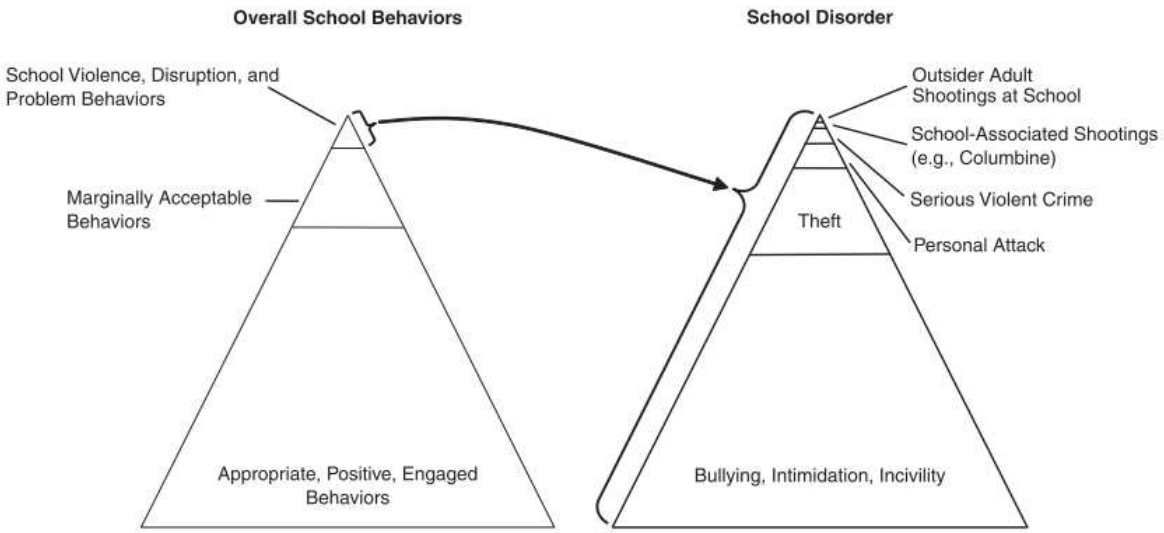


Figure 9. Schematic of the Structure of the School Safety Research. Reproduced with permission from Mayer and Furlong (2010, p. 17).

Security from Outside Threats. Target hardening, the practice of improving safety by improving the physical security of school buildings and grounds, involves identifying and implementing physical changes to the layout, sight lines, fittings, and structure of the physical grounds of the school. Both an FBI study into school safety (O'Toole, 2000) and a U.S. Secret Service / U.S. Department of Education study (Vossekuil, Fein, Reddy, Borum, & Modzeleski, 2002) called for target hardening across schools in the U.S. as a matter of urgency. In order to protect students from outside threats, particularly outside adult shootings and school associated shootings, the consensus perspective was a *threat assessment* approach, as described in Cornell and Williams (2012). The threat assessment approach involves early identification of threats,

followed by a progressively escalating response, constantly evaluating the level of threat and responding appropriately (see Borum, Cornell, Modzeleski, & Jimerson, 2010; O'Toole, 2000; Vossekuil et al., 2002).

Earlier findings on peer victimization as a source of suicidal or self-harm behavior (Jutengren et al., 2011; O'Connor et al., 2010) have elevated urgency when we consider security from outside threats. Vossekuil et al. (2002) found that suicidal ideation, dreams and fantasies of murder/suicide are embedded in over three-fourths (78%) of school shootings. Miller (2012) described a number of protective factors against youth suicide and self-harm, many of which are recommended to be built into anti-bullying practice: building suicide and self-harm warning signs and reporting protocols into student and teacher anti-bullying and upstander training and having school staff trained in suicide intervention practice.

School Community

In their mixed methods research based in Israel, Astor et al. (2009) found that low violence schools have a strong outward facing perspective, where teachers and school leaders saw their role as preparing young people for emergence into the community, and purposefully provided a caring nurturing place to teach their students about the world. This was contrasted with high violence schools, where leaders and teachers had more of an inward focus, with academic goals having unrivaled primary importance. School leadership – particularly the leadership style exhibited by the school principal – is clearly a critical component of positive school climate, both in terms of safety and community.

Family Support. Parents are important influencers on their children's behavior in school. Nickerson et al. (2008) found that attachment to mothers was a strong predictor of upstander behavior in middle school aged children. In their analysis of the effect of parental

involvement among Hispanic and Latino middle school students in South Florida, Adamski, Fraser, and Peiro (2013) found that parental involvement in school life was highly predictive of attitudes toward learning. Their analysis included comparative factor analysis with both home and classroom environments, showing that, while the home environment was most predictive of student *attitudes*, the classroom environment was most predictive of student *achievement*. A deeper, qualitative exploration of the relationships between minority student families and their schools was conducted by Lawson and Alameda-Lawson (2011). In their action research case study, the researchers examined and supported Latino families as they entered and maintained their relationship with a school community in a low income neighborhood, concluding that action researchers and families themselves have the social capacity to influence educational success for their children, if they organize and learn from the experiences of others. The implication of these findings is that while family influences are essential to preparing students to learn, learning is primarily driven by teachers, delivered through positive learning environments. Parents and teachers must collaborate to drive positive outcomes for their students.

One important area of research is focused on homework. In their research synthesis, Patall, Cooper, and Robinson (2008) first synthesized 14 studies of homework training for parents, showing that training and encouragement of parents leads to higher levels of homework completion and fewer homework problems. They then conducted a synthesis of samples from 20 additional studies, showing that parental involvement in homework is positively correlated with academic success in elementary school and high school, but negatively associated with academic success in middle schools. Parental involvement in mathematics homework showed a negative association with success in mathematics across the board. Taken together, these findings suggest

a sophisticated, evidence-based program of engagement between teachers and families, in order to create opportunities for student success.

Beyond direct family support, Chan et al. (2013) examined the effect of school-based mentoring programs (in this case, an onsite Big Brothers Big Sisters program) on student engagement and relationships with teachers and families, and ultimately on academic outcomes. The researchers concluded that these programs have a significant effect on relationships with parents and teachers, and were associated with improved perceptions of school climate. This finding implies that appropriate school-based adults other than families and classroom teachers have an important place in student success.

Supportive School Staff. In a qualitative phase of their mixed methods study with middle school students in England, Pryce and Frederickson (2013) found that students had positive feelings about their school when school staff were consistently viewed as listening and being attentive to student needs, and negatively about their school when they felt ignored, or when they had an inconsistent perception of staff attentiveness. Students also projected these positive and negative perceptions into expectations about how their school would react to incidents of bullying and acting out behavior. In her analysis of data from the National Institutes of Health, O'Connor (2010) found that teacher-child relationships generally deteriorated from first through fifth grade, although high quality, supportive classroom climates had a mitigating effect, slowing the deterioration in teacher-child relationships.

Hand (2009), in her observational research in middle school mathematics classrooms, found that teachers sometimes created conflict and oppositional behavior in their students by their demeanor and practices. She concluded that weak mathematical engagement, and poor teaching practices were an important cause of oppositional behavior among students, particularly

among students prone to math anxiety and poor math performance. Outside of the mathematics classroom, Danielsen, Wiium, Wilhelmsen, and Wold (2010) conducted a structured equation modeling analysis of data on Norwegian 13 year olds, concluding that teacher support varied significantly between classrooms, and that it was predictive of improved academic initiative and motivation. In the same study, researchers found that perception of peer support also had a significant (but slightly lower) effect on academic initiative. Givens Rolland (2012), in her meta-analysis of classroom goal structures, found that teacher socio-emotional and instructional support was positively associated with academic success. Martin (2008) showed that multilevel motivational interventions by teachers and school leaders can increase engagement, motivation, and ultimately academic success, when implemented with fidelity. Essentially, teacher support and engagement has a direct impact on learning motivation, and peer support and engagement has a secondary, direct impact on learning motivation.

This learning appears even more important when teaching minority populations, such as the Maori peoples of New Zealand. Bishop, Ladwig, and Berryman (2013) showed how direct relationships between teachers and learners were embedded in cultural norms among indigenous peoples, concluding that not only are teacher-child relationships important for building social cohesion (as in *Western* cultures), but also are foundational to indigenous pedagogies, with the spirit of the wise one or guru nurturing her disciplines toward their own adulthood and wisdom. This has important implications for connective instructional strategies, particularly in schools that serve indigenous populations in the U.S.

In their meta-analysis of data from 99 studies examining teacher-student relationships, Roorda, Koomen, Spilt, and Oort (2011) concluded that positive teacher-student relationships were associated with both student connectedness and achievement; achievement association was

weak (although significant). Negative teacher-student relationships in lower grades had the most significant negative impact on both connectedness and achievement in later grades. In addition, Rudasill, Reio, Stipanovic, and Taylor (2010) found that negative teacher-student relationship in early grades predicted risky behavior in later grades. The implications of these findings are that teachers, especially in the lower grades, need to set the tone for improved outcomes in later grades by establishing close positive relationships with all students, particularly those who might be at risk for later problems at school.

In addition to shifts in teacher-student relationships over several grades, recent research by Turner, Gray, Anderman, Dawson, and Anderman (2013) showed that, while teacher and student perspectives might differ at the beginning of a particular school year, they tend to converge toward the end of the school year. The obvious implication of this is that students and teachers adapt to each other through the year, as they get to know each other. This finding strengthens the value of deliberate action to provide a supportive nurturing teacher-student relationships for all students, and implies that researchers must take care to observe these shifts in perception during the school year.

Student Cohesiveness. In their integrative review, Martin and Dowson (2009), unified a variety of theories of achievement (for example, attribution theory, expectancy-value theory, self-determination theory) using relatedness as their framework. This work demonstrated the importance of relatedness and relationship as drivers of student motivation and success. The authors proposed an action-oriented framework, based on the theory that improving student-student and teacher-student relationships would drive motivation and thereby achievement for all students. Ciani, Middleton, Summers, and Sheldon (2010) found that support for student feelings of community in the classroom (along with support for autonomy) had a mediating

effect, shielding students from the potential negative impact of modern performance-oriented classrooms, providing a mechanism to offset the impact of standardized testing, performance anxiety, and other externally imposed goal structures.

Student relatedness and *positive* peer engagement are associated with lower degrees of violence and improved academic outcomes in school. In their study of Australian 9th graders, Chapman, Buckley, Sheehan, Shochet, and Romaniuk (2011) found that school connectedness had a protective effect on violence and multiple risk-taking behaviors that lead to injury. Gini, Pozzoli, Borghi, and Franzoni (2008) conducted an experimental study to understand students' perceptions of safety, based on bystander behavior in scenarios of direct and indirect bullying and aggression. The study showed that students had positive, generally protective feelings toward victims of bullying and negative feelings toward aggressors, but, that the attitudes of other bystanders had a significant effect on perceptions of safety, even within the same situations of aggression. The implications of this research are significant: upstander action is critical in order to improve the perception of safety in a school.

In a comparative experimental study conducted with middle school students in suburban Texas, Loukas, Ripperger-Suhler, and Herrera (2012) concluded that the driver of school connectedness was more likely to be adolescent problems (depression, negative relationships, conduct issues) than incidents of peer victimization. One implication of this finding was that attention should be given to providing students with social-emotional coping strategies in early adolescence, to help deal with normal adolescent problems. Students better prepared to cope with the *drama* of adolescence were more likely to feel connected to their school, and therefore tend to engage in fewer negative behaviors. Further, in their longitudinal study, Niehaus, Rudasill, and Rakes (2012) found that 6th grade students, particularly boys, felt that levels of

support and connectedness with school deteriorated during the school year, but that those students whose connectedness deteriorated less (or increased) during the year, had higher academic achievement.

In their examination of the transition from 7th to 8th grade, Wang and Holcombe (2010) found that 7th grade students' sense of participation in and belonging to their school, and their sense of self-regulation were positively associated with these same dimensions of school climate in 8th grade, and with their 8th grade levels of academic achievement. Positive and negative experiences at school do carry from year to year, even though leaders, teachers, and classrooms might be very different. This finding might explain why school climate (whether positive or negative) tends to persist in a school until some significant and visible change in direction and leadership occurs (Astor et al., 2009).

Respect for Diversity. In their mixed methods analysis of atypical violence schools, Astor et al. (2009) found that the low violence schools in Israel had significantly higher levels of respect for diverse populations when compared with the high violence schools. Given the ethnic homogeneity of all the schools in their study (designated as either Jewish or Arab schools), diversity referred specifically to disabled students. While the low violence and high violence schools were similarly homogeneous, the low violence schools were observed to have higher degrees of messaging about the importance of accepting differences, both in terms of ethnic and disability status:

> In numerous wall decorations the idea of inclusion is expressed. For instance, in a picture of Noah's Ark, those approaching it included Jewish and Arab children and children using wheelchairs. (Astor et al., 2009, p. 449)

Monsen, Ewing, and Kwoka (2013) studied the impact of teachers' attitudes to inclusion, specifically with respect to inclusion of students with special needs in the mainstream classroom, concluding that teacher attitudes to inclusion were closely related to overall classroom climate, engagement, and the perception of students of teacher support and engagement. The more teachers have a focus on justice and equity, the higher level of trust students felt toward them, and the stronger relationships that formed among them. Peter and Dalbert (2010) explored German secondary school students' perspectives on what they called Belief in a Just World (BJW), connecting this belief to school climate and degree of teacher focus on justice and equity. They concluded that higher justice teachers created more just and equitable classrooms, leading directly to improved classroom climate, particularly in diverse settings.

Goldsmith (2010) explored achievement scores among racially diverse student populations, using persistently segregated schools (predominantly white or black student populations) as context, concluding that normative effects (where students try to be like their peers) and frog pond effects (where students try to stand out from their peers) interact to explain some of the differences in outcomes between these segregated schools. This finding has implications for engagement of diverse populations, particularly in high diversity school settings.

Learning Environment

In their article describing the Individualizing Student Instruction (ISI) classroom observation system, Connor et al. (2009) demonstrated a clear understanding of the major components of a classroom, with particular focus on how teacher characteristics (warmth/responsiveness, content area knowledge, control/discipline, understanding of student needs) interacted with student characteristics (language, literacy, self-regulation, social skills, home support), in the context of the instructional environment (content, level, groupings, teacher

or student directed, duration/time on task), to create student outcomes. They pointed out that this kind of analysis provided a picture of the context in which students learn, and that levels of individual student emotional and cognitive engagement are likely just as important in driving student outcomes as are the dimensions measured by the ISI observation system. They did, however, conclude that to create positive student outcomes, classrooms should intentionally include understanding of individual differences in learning style, employ formative assessment processes aligned to goals and expectations, and be responsive to each student's cognitive, behavioral, and social-emotional needs at each point in time during the class. In a later project, Connor et al. (2014) used an additional, more global observational tool, combined with the ISI framework, and concluded that global characteristics of the learning environment and time spent on specific learning tasks together led to improvement in learning, whereas neither did so on their own. This implies that both content specific activities and environmental concerns must be considered in order to improve classroom learning.

High Expectations. Teachers establish goals and expectations for their classrooms, and both teachers and learners use extrinsic and intrinsic motivation to drive success against these expectations. A recent experimental, randomized study in New Zealand showed a direct and significant intervention effect when 46 math teachers were provided with a series of professional development workshops with observational follow-up mentoring, modeling the practices of high expectancy teachers. Significant improvement in math achievement was observed over a one year intervention period, when compared to the 41 math teachers in the control condition (Rubie-Davies, Peterson, Sibley, & Rosenthal, 2014).

Givens Rolland (2012), in her meta-analysis of classroom goal structures and student achievement, showed that extrinsic, outside imposed goals can have a stifling effect on student

self-motivation, potentially leading to lower achievement. The implication of this finding may be that teachers must leverage both intrinsic and extrinsic motivation – set high expectations and continue to drive for internal and external goals – in order to best motivate students to achieve. In addition, research has shown the importance of teachers starting the year well and sustaining their positive demeanor and emotional closeness to their students throughout the school year. In their extensive modeling study of elementary schools in Portland, OR, Corpus, McClintic-Gilbert, and Hayenga (2009) found that while intrinsic motivation was a powerful motivational force, extrinsic motivation was almost as effective and did not significantly impact classroom grades. This study also showed that intrinsic motivation generally deteriorates with age, particularly in the period from 3rd through 8th grade, while extrinsic motivation was relatively stable across these time period. This implies a motivational approach which aligns extrinsic motivation with intrinsic motivation for maximum effect.

Mainhard, Brekelmans, Brok, and Wubbels (2011) showed that classroom climates were highly stable, and have a slightly U-shaped profile during the school year. The implication of this finding is that teachers need to establish emotional closeness from the beginning of the school year. In an analysis of a similar question in U.S. schools, Curby, Rimm-Kaufman, and Abry (2013) concluded that higher emotional support at the start of the year was associated with higher instructional support later in the year, and that higher instructional support earlier in the year led to a need for higher emotional support later in the year. Interestingly, teachers trained in responsive classroom (RC) techniques showed higher emotional support and higher instructional support levels (from student perspectives) throughout the year.

In their hierarchical linear modeling analysis of perceptions of learning environments among Australian middle school girls, Spearman and Watt (2013) found that student perceptions

of their learning environment (as measured through self-report questionnaires) differed significantly from the *reality* of their learning environment (as measured by expert observation). This is highly significant, because the students' perception of their learning environment was *more closely associated with academic outcomes than the reality measure*, implying that when it comes to expectancy-value theory and school climate, *perception is more important than reality*.

McKown and Weinstein (2008) examined teacher expectations, using an ethnic achievement gap lens, and showed that, in classrooms where students reported bias, teacher expectations of African American and Hispanic students were lower than their white or Asian American peers, even when students had similar records of prior achievement. The researchers concluded that this unconscious teacher expectancy bias accounted for a moderate but significant component of the ethic achievement gap in the study classrooms. Teacher expectations are powerful.

Effective Teaching. Marzano (Marzano, Marzano, & Pickering, 2003), described the role that teachers play in creating a positive learning environment, as follows,

> Effective teaching and learning cannot take place in a poorly managed classroom. If students are disorderly and disrespectful, and no apparent rules and procedures guide behavior, chaos becomes the norm. Teachers struggle to teach, and students most likely learn much less than they should. In contrast, well-managed classrooms provide an environment in which teaching and learning can flourish. But a well-managed classroom doesn't just appear out of nowhere. It takes a good deal of effort to create – and the person who is most responsible for creating it is the teacher. (p. 1)

Marzano provided a set of research-based guidelines for teachers. For example, a set of rules and procedures were suggested for transitions and interruptions, and for distribution and

collection of shared materials and resources during the school day. These rules and procedures should be rehearsed early in the school year, and once established followed consistently throughout the year. Marzano et al. (2003) also raised up the need for mindfulness in the classroom (pp. 65-75), including the concept of *withitness* – that is, teaching with clarity of purpose combined with a high degree of attention to individual student needs and a keen understanding of the demeanor of the whole group.

Diverse Learning Styles. Several research studies in this review examined teaching practices which integrate academic and social learning, providing a variety of learning experiences for students. Brock, Nishida, Chiong, Grimm, and Rimm-Kaufman (2008) examined the Responsive Classroom (RC) approach in their longitudinal study of student perspectives on their classroom learning environment. The RC approach is constructed on a set of principles and practices including a class morning meeting, collaborative development of a set of classroom rules and consequences, a particular classroom organization designed to maximize learning and interaction between students, guided discovery learning, and academic choice. This approach is deliberately designed to foster intrinsic motivation and autonomy among RC learners. The study found that students in classrooms with more variety of RC practices had better academic outcomes, less behavioral issues, and improved perception of their school. A later randomized study on the RC classroom (Rimm-Kaufman et al., 2014) showed that while RC practices did not show a significant improvement in student achievement in math and reading overall, it did demonstrate disproportionate improvement for those students whose initial math achievement lagged their peers. This second study also examined the effects of implementation fidelity in both the treatment and control groups, finding that some RC practices were already being used in control classrooms, and some treatment classrooms did not implement some RC

practices with high fidelity. The researchers concluded that fidelity of implementation mediated their results.

In their experimental study of five elementary schools, Reis, McCoach, Little, Muller, and Kaniskan (2010) found that the School-wide Enrichment Model – Reading (SEM-R) was effective at improving reading skills compared to control schools. This is important in terms of diverse teaching and learning styles, because the SEM-R model employs a variety of learning contexts, and enriching environments. This broad finding is similar to findings by Reznitskaya et al. (2012), that while an inquiry based learning pedagogy (Philosophy for Children, similar to Socratic method) was effective in increasing 5th grade students' argumentation skills during class, (and presumably was more engaging for students), this intervention did not significantly increase argumentation skills in a post-intervention assessment when compared to regular (control) classrooms.

Problem Based Learning (PBL) is another teaching and learning model that could be employed to provide a diverse learning environment in K-12 classrooms. Wirkala and Kuhn (2011) investigated the efficacy of PBL in a randomized controlled trial, with three sets of middle school classrooms being compared: a lecture based, non-PBL classroom, a group-PBL classroom, and an individual PBL classroom. Both PBL classrooms succeeded in increasing 9-week post implementation achievement versus the lecture classroom, and there was some variability among individual students between the group PBL and individual PBL classrooms. This latter finding suggests that certain students benefit from the group classroom and others from the individual PBL classroom, although all students benefited from PBL structures when compared to the lecture classroom.

Challenging Curriculum. In his significant mixed methods study, Cooper (2013) investigated practices for effective student engagement by examining three dimensions of teaching practice. First, mirroring the work of Martin and Dowson (2009), he looked at connective instruction (purposeful connection of content, instruction, and relationships to each individual student's emerging personal identity), showing these teaching practices to be by far the most engaging teaching practices in the study. Cooper (2013) also looked at academic rigor, finding challenging work to be critical to engagement, provided it is properly connected to student identity. In situations where challenging content was taught without being connected properly to individual student identity, it was seen to have a weaker effect on engagement. Finally, he examined the practice of lively teaching, using fun, active, experiential learning processes. Once again, this practice led to engagement, but was amplified by combination with connective instructional practices. Plank and Condliffe (2013) studied 2nd and 3rd grade classrooms in Baltimore to examine the quality of the learning environment in classrooms under pressure to improve standardized test scores. They found that in classrooms under the most pressure to improve, focus shifted away from quality teaching interactions into an urgent focus on direct instruction and test preparation drills, in an attempt to directly improve test scores. These differences were remarkable during the early spring as students were prepared for standardized testing, and essentially disappeared by May, when test pressure had subsided in the schools. The implication of this is that *teachers knew* they were engaging in a negative activity, in terms of actual learning outcomes, but executed the plan anyway, returning to their typical classroom environment as soon as they were able to.

In their small scale quantitative analysis of emotional affect in the classroom, Goetz, Lüdtke, Nett, Keller, and Lipnevich (2013) found that teachers who exhibit a supportive

presentation style, that is, pedagogies focused on explicitly helping students to achieve mastery, tended to create more positive emotional states in their students, particularly when dealing with complicated subject matter. They contrasted this teaching style with excessively demanding presentation styles, which left students feeling unsupported and anxious. This finding was not directly associated with the complexity of the material being covered (although it did encompass multiple domains of content), but implied that difficult material is viewed by students in a negative light when taught in a demanding way, and the same material can be viewed by students in a positive light when taught in a more supportive way.

Creativity and Higher Order Thinking

New Bloom (Anderson & Krathwohl, 2001) provided a clear definition of higher order thinking, as those cognitive processes in the top three levels of the taxonomy – analyzing and evaluating information, and creating new products and knowledge based on that analysis and evaluation. As illustrated in Figure 10, the highest level of higher order thinking involves the creative act.

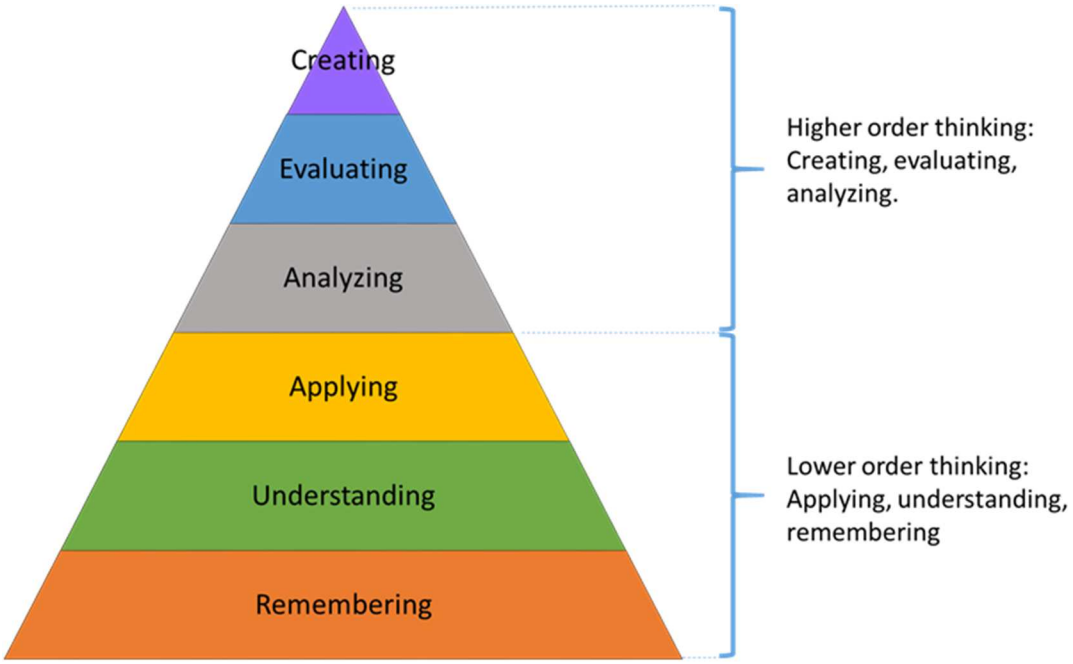


Figure 10. New Bloom's Taxonomy Categorized into Lower and Higher Order Thinking.

Defining Creativity. Creativity has been defined in a number of ways since the serious study of creativity started in the 1950s. Most contemporary definitions of creativity (Kaufman & Sternberg, 2007) include three components: First, that the product of creativity is novel in some meaningful way; second, the product is of high quality, as defined within its context; and, third, that the product is useful, again, within its context. For the purposes of this dissertation, the following working definition of creativity was used: *Creativity is the ability to produce work in any domain that is novel, exhibits high quality, and is useful to its domain.*

Csikszentmihalyi's Flow Theory of Creativity. Amongst the many theories of creativity (for an excellent review of theories of creativity, see Kaufman & Sternberg, 2010), Csikszentmihalyi (1996) provided a germinal systems perspective on creativity, with an emphasis on the creative environment and the interaction of influencing factors on the creative act. The primary factors considered in this model are the creative *domain*, *field*, and *person*.

The domain is the area of expertise in which the creativity occurs – art, math, engineering, scientific discovery, business, etc. The field is primarily defined in terms of the *gatekeepers*, the people operating in the domain who judge and guide the creative act. Clearly the domain and field are tightly connected – the domain specifies who are the gatekeepers, and the gatekeepers are, in turn, highly influential on the domain. Operating within the domain/field continuum, are the persons engaged in creativity. In essence, the theory suggested that individuals begin with certain inate creative abilities, but the combined influence of the domain of work and the influence of gatekeepers, provided a model for improvement in creativity. This theory further suggested that the development of individual creativity is an evolutionary learning process, where multiple creative experiences are required over time, in a supportive and engaging creative learning environment.

In creating an *environment* for teaching and learning higher order thinking, it is important to first understand how children create things. Lassig (2013) used a grounded theory approach to explore the ways Australian adolescents used creativity in production, identifying four major paths to generativity: adaptation of existing ideas into a new idea; transfer of ideas from one domain into another; synthesis of two or more ideas into a single new idea; and, genesis of new ideas, from scratch. The major educational recommendation from this study was to bring metacognition into learning. Explaining how creativity works to students, in her view, supported and empowered students to be creative in their own practices, as well as providing an opportunity to highlight and reinforce the value of creativity in school and in life.

Creative Focus. The core idea in Csikszentmihalyi's theory (Csikszentmihalyi, 1996) was that, in order to be optimally productive, the creative person needed to enter a state of focus called a *flow* state. In this model, creative flow began with what is *fun* for the creator, what the

creator *loved* to do. Much of Csikszentmihalyi's ongoing research over 40 years in the domain (for example, Csikszentmihalyi, 1975, 1988, 1990, 1996), focused on interviews with creative people, and clearly supported the idea that this sentiment is universal, across all domains of work, that an individual can only be creative in work she loves, where she finds fun. The ability to enter flow state is, however, a *learned, practiced* ability. In order to be creative, as in any complex endeavor, the creator needs both experience and practice focused on the creative act. Through practice, the creator becomes more adept at the creative process, and becomes more able to produce creative products – Csikszentmihalyi (1996) used the word *autotelic* to describe the extent to which creativity became an instinctive practice among those who practice it repeatedly over time (p. 113). While we do not expect students to forget to eat, drink, or sleep, in order to perform creative acts, focus and flow are essential components of the environment for creativity (Csikszentmihalyi, 1990, 1996).

Traditional thinking about creativity implies that time is required to execute creative processes, more precisely, that time pressure inhibits creativity (for example, Amabile, Hadley, & Kramer, 2002; Mainemelis, 2002; Wallach & Kogan, 1965), making it very difficult to integrate creativity into 40 minute class periods in a public elementary school. In their study on time pressures and creativity, Antes and Mumford (2009) found that while time pressure can have a suppressive effect on creativity overall, it can have a positive effect on particular components of creative process and problem solving, particularly if properly framed (in either a positive light for generative steps or a negative light for evaluative steps) by the facilitator. Other inhibitors of creativity included classroom and group cohesion. In a mixed methods study of group problem solving in Californian high school mathematics classrooms Chiu (2008) found that rudeness and over-critical evaluations of peer ideas led directly to reductions in creative

generativity in subsequent creative tasks. Students who behaved in friendly, supportive ways helped create higher quantity and quality of ideas in subsequent interactions.

The emotional effect of students preparing for a creative task, as well as the emotional effect of the creative act upon the students must be considered in addition to time pressures in a school classroom. Students appeared at the start of class carrying the effect of their prior class period, of the drama of their young lives, or of their family and peer relationships, and their feelings about the class they are about to enter. In his discussion of the interactions of emotions and creativity in classroom contexts, Newton (2013) proposed a framework for teachers to consider the management of these complex emotions in preparing for creativity tasks. He suggested that teachers should first frame the creativity tasks in appropriate ways, for example, through direct knowledge of what constitutes appropriate and novel within the domain being studied. Holding up excellent examples may, however, hinder creative generation or cause emotional downturns in students, since they might feel they could never achieve such excellence. Having carefully framed the task, teachers should facilitate the creative process consciously, providing support while simultaneously creating an environment where creative flow can occur in as many students as possible. The goal is not only to complete the assignment, but also to stimulate in students the sense of joy that emerges from the creative act. This is both a tactical need within this particular class period, and a strategic move, in preparing the emotional expectations of students for the next class.

Chung (2013) compared levels of creativity (as measured by divergent thinking tasks) between habitual Taiwanese players of a particular role playing game (similar to *Dungeons and Dragons*) and non-players, finding a significant association between game play and divergent thinking and improvisation. The implication of this is practice is that learners will become more

creative if they are exposed to repeated creative tasks across their school experience. In her exploration of student perspectives on the infusion of creative elements into Chinese science classes, Cheng (2011) interviewed students who had experienced a significant shift in science classrooms for a period of about four months, during which time instructors deliberately infused creative activities and experiences into their science classrooms. The most positive responses from Chinese students included the playful and active learning environment created in this new learning style.

Creative Freedom. In organizational development as well as in school management, over the centuries, firms and schools have established systems of control, designed to ensure order and consistency in the system. In situations where all knowledge is existent, and the job of the learner is to master existing knowledge, these systems might be applicable. However, when examining creativity, it is clear that bureaucratic rules and procedures – the bread and butter of classroom management – may have a suppressive or dampening effect on creative processes in the classroom. In the organizational development world, bureaucratic processes have the deliberate purpose of sustaining established patterns of behavior and supporting the hierarchy, thereby rejecting unfamiliar creative ideas (Williams & Yang, 2010).

In order to be creative, humans need to overcome organizational or cultural barriers to creativity and overcome a variety of personal barriers to creativity. One of these is known as fixation (or premature closure). This is particularly relevant in school settings, because overcoming fixation requires an environment in which students can use their knowledge to generate solutions without getting stuck on the *first right* answer. Agogué, Poirel, Pineau, Houdé, and Cassotti (2014) explored this issue with elementary, high school, and college students, using a mix of quantitative and qualitative approaches to examine the degree to which

different age student populations, and differently trained student populations were differentially inhibited by fixation. The research showed that both *lack of knowledge* about potential solutions and *excess of knowledge* about potential solutions caused fixation. The implication of this for classroom settings is that creative problem solving should *precede* knowledge transmission. Before teachers reveal existing knowledge or solutions, students should be engaged in creative problem solving, teachers should then facilitate the process of transferring new solutions into new knowledge for the students, allowing students to feel that they have *discovered* the solution, even if the discovered solution was already known to the teacher.

Ciani et al. (2010) found that support for autonomy (combined with student feelings of community) in the classroom had a buffering effect, shielding students from the potential negative impact of modern performance-oriented classrooms. The implication of this finding is that positive classroom environments, and particularly support for student autonomy, can be used both as direct instructional tools and as a mechanism to offset the impact of standardized testing, performance anxiety, and other externally imposed goal structures. Davies et al. (2013) explored the learning environment for creativity, finding that freedom to take risks combined with support for learning, delivered in a loosely structured plan worked best to engage primary students in learning to use creative problem solving techniques.

Abrami et al. (2008) conducted a meta-analysis of 117 studies into *critical thinking*, another important form of higher order thinking, and, showed that one cannot assume that critical thinking will be delivered as part of content oriented classrooms. Mathews and Lowe (2011) went further, detailing exactly which practices served to suppress critical thinking and which served to foster such dispositions. Specifically, autonomy and independence supported by figures in authority were found to be essential in development of higher order thinking, while

deference to authority, memorization of predefined knowledge, and harsh response to initial ideas served to inhibit higher order thinking. Schools should engage in purposeful definition, professional development, and curriculum and lesson plans designed to teach critical and creative thinking. Higher order learning does not happen by accident, particularly in times when test performance is the primary goal of education.

In her interviews with Chinese students undergoing infusion of creativity into their science classrooms, Cheng (2011) described students feeling incredible freedom to express their ideas and to support the ideas of their classmates. While some students struggled with the sudden perceived lack of structure in the classroom, they viewed this change as positive for their overall learning.

Creative Challenge. In his early work on flow in creativity, Csikszentmihalyi (1996) described the importance of creative work being challenging, but not too challenging, as follows,

> The pursuit of a creative problem is rarely easy. In fact, in order to be enjoyable it should be hard, and of course so it is, almost by definition. It is never easy to break new ground, to venture into the unknown. When one starts out, the difficulties may seem almost overwhelming . . . . It is impossible to accomplish something that is truly new and worthwhile without struggling with it. It isn't just in sports that the saying 'no pain, no gain' applies. (p. 116-117)

In their comprehensive review of creative education in Scotland, Davies et al. (2013) showed that creative projects need to be both complicated and authentic to real world experience, in order to motivate deep creativity in students. Students must believe that the project does not have one pre-defined solution, must believe that they have some autonomy in creating their solution, and that their solution will be valued both inside and outside of the school. One

example of this was a local environmental project involving elementary students in community remediation work.

In Cheng (2011), students noticed the complexity and scope of the challenges they were given, sometimes finding them obtuse or wild, and missing the scaffolding typically provided by the instructor. This was perceived by Chinese students as one of the least positive aspects of the creative classroom, and yet, students also saw this experience as one of the most powerful learning experiences from the study.

Support for Creativity. DiLiello, Houghton, and Dawley (2011) examined the effects of perceived support for creativity on creative self-efficacy and self-perception of creativity, and found that supervisors and work-group members are the primary moderators of the creativity gap between diverse employees. While this has not been well researched in school settings, the implication is that teachers (supervisors) and work-group members (student peers) should be trained to support creativity in order to improve the creative climate in a school.

Kim (2008) concluded some teachers treat creativity like a disability. Students who exhibited creativity in the classroom, whether inherent or learned, often co-exhibited behaviors which were characterized by teachers as nonconformist and untraditional. He recommended several potential solutions to this issue, including teacher training on diversity and support for creativity, and the use of models of giftedness education for creative underachievers, such as Renzulli (1992). In order to specifically support creativity, teachers might develop classroom goals and incentives to help motivate students to be more creative. Peng, Cherng, Chen, and Lin (2013) found that mastery/acceptance goals had the most significant impact on student motivation toward creativity, with performance/avoidance goals having no significant impact. Student motivation is important in driving creativity in the classroom, and both recognition of

creative acts and development of goals that reinforce the value of creativity can help develop this motivation.

In his professional text on creativity, Mark Runco (2007) described how teachers can adopt the role of mentor, taking advantage of their role as role models for many students, by directly exhibiting creative behaviors. More often than not, however, teachers take the role of squelchers, using language and behaviors that subconsciously suppress creativity. Support for creativity in the classroom must be both transparent and consistent for it to be effective.

Defining a School Climate to Encourage Creativity

In their study of *creative climate* in British middle school design and technology classrooms, McLellan and Nicholl (2013) used Ekvall & Isaksen's creative climate model (for a detailed description, see Isaksen & Isaksen, 2010), applying it to classrooms and using mixed methods to assess the fit between student perceptions of the climate in their classroom and their adaptation of the organizational creativity model, and to deeply analyze student perceptions of the climate in their design and technology classrooms. They concluded, first, that the model had some utility as a framework for assessing classrooms for creative climate, but that more work was needed to test the psychometric properties of the measure in classroom settings. Their second, and more significant conclusion, was that students generally did not feel that their school climate was conducive to engaging in creativity. For example, students felt that they were not being challenged to use their creativity enough, had insufficient freedom to find their own solution paths, etc. It was clear that some of the components of creative climate were missing, even from a design and technology classroom, which might under normal circumstances be seen as a creative subject area. The implications of these findings are significant, and help establish

the importance of research focused on learning environments for creativity and higher order thinking.

Rubenstein, McCoach, and Siegle (2013) developed and tested an instrument to measure teachers' perceptions of the factors which would allow them to teach creativity in their classrooms. As expected, teachers felt hindered by their perceptions of lack of time and lack of support. In this study teachers clearly saw the value of teaching for creativity, had high levels of self-efficacy, and felt that their students possessed the potential to learn to be more creative. The authors speculate that this could be because of their convenience sample, or, it could simply be that teachers generally agree that creativity is important, but do not feel that they are allowed to develop creativity in their students.

By contrast, Chinese students in Hong Kong were engaged in a large scale long term study to deliberately infuse creativity into science classrooms. In her analysis of student responses, Cheng (2011) found positive responses from her participants in terms of learning outcomes, but simultaneously observed altered responses to creative activities in general. Chinese students, most of whom had never experienced any creative learning activities before this study, found the activities difficult and overwhelming, and to a certain extent, found them to be a waste of their time, when they could be learning from their instructor. Cheng concluded that infusion of creativity skills into Chinese science classrooms was probably not the best approach, proposing instead, that Chinese schools should have created separate curriculum for creativity. Overall, it seems that Chinese students valued creativity less after this experience: they did, however, feel that the creative classroom activities they experienced were highly effective for learning science, more effective than their previous didactic learning experience. She was clear that the challenge in her project was that her model was based on *Western*

creativity literature, but applied in an *Eastern* cultural context. This article raises a fascinating insight, that the cultural dimensions of both the existing learning landscape and culture, and the model for teaching and learning creativity being implemented must be carefully examined to ensure cultural fit. Context is *important*: What works in one place may not work in another.

Despite this finding about cultural fit of models, for some, multicultural experiences are strong predictors of readiness for creative development. Leung and Chiu (2008) experimented with *creative potential* and *openness to experience* in undergraduate students, finding that those with extensive multicultural experience in their background had significantly higher creative potential only when they were also highly open to experience. Those who had multicultural experiences but were not open to experience exhibited lower creative potential. The implications of this include a focus on teaching students to be open to experience, potentially by using generative practices and by supporting creativity throughout school learning and across the curriculum, rather than using unique creative multicultural experiences in an attempt to stimulate creative response.

The Proposed Climate4Creativity Model

School is a Safe Place. A school designed with a climate for creativity would begin by being a safe place. A place where students can both feel safe and actually be safe. This would manifest itself in three domains of safety: 1) *Safety from inside threats* such as bullying, violence and aggression; 2) Creation of a *safe built environment*, where there are no hot spots or danger areas on school grounds; and, 3) *Security from outside threats*, where school leaders proactively protect students from incursion by unauthorized persons.

School is a Caring Community. A school designed with a climate for creativity would be a caring community of teaching and learning, nurturing children and adults in an environment

of mutual kindness and respect. This would manifest itself in four domains of community: 1) *Strong family support*, including collaboration between school professionals and family members, direct support for student learning, and positive parental activism; 2) A consistently *supportive school staff* would teach and mentor and guide students as they learn, providing appropriate levels of academic and emotional support to enable students to develop positive habits and practices; 3) A *cohesive student body* would enable students to help each other and form a micro-community within the overall school community, providing a protective and supportive environment for students to learn; and, 4) such a community would exhibit *respect for diversity* in terms of race, gender, sexual orientation, physical attributes, socio-economic status, opinion, etc.

School is a Place for Learning. In the context of a safe, caring school community, every classroom would be designed as an engaging learning environment. This classroom learning environment would manifest itself in four domains: 1) Teachers and students would have *high expectations* of each other, and would be open and honest about these expectations; 2) *Teachers would be effective* in all classrooms, managing their time well and providing clear instruction to their students; 3) Teachers would respect and engage a *variety of learning styles*, activities and assignments would be diverse, and students would have opportunities to work individually and in team as they learn; and, 4) There would be a *challenging curriculum*, pushing students to perform at the level of their potential.

School is a Place for Creativity. In a school designed to exhibit a school climate for creativity, every classroom would support creative practices. Classroom support for creativity would manifest in four major domains: 1) *Fun & Flow*: Students would have fun while learning and would experience creative flow most days at school; 2) *Creative Freedom*: Students would

be encouraged to reframe problems for creativity and to create their own solutions to classroom assignments; 3) *Creative challenges* would encourage students to think deeply about the topics they are studying, and would ensure that creativity was required for each assignment, rather than being a separate activity for some assignments; and, 4) teachers and school professionals would *support and encourage creative practices* across the curriculum.

An Ideal Creative School. In her grounded theory research, Kangas (2010) explored Finnish students' perspectives on the ideal school, developing a grounded theory of what she calls a *Broadening and Empowering Learning Environment (BELE)*, as follows:

> This model school, called a Broadening and Empowering Learning Environment, is designed to contribute to children's physical, educational, cultural and socio-emotional well-being, offers opportunities for fantasy and innovation, and employs creative and sports-based learning methods, among others, in both formal and informal settings. The study showed that children, as educational stakeholders, are well aware of the potential of modern schools and of the different aspects that would enhance their ability to learn and their satisfaction with schooling. (Kangas, 2010, p. 1)

The BELE developed by Kangas, on behalf of the students in her study, is structured along four dimensions, which appear to connect well with the themes extracted from the research literature on student perspectives in this space. Table 3 provides an outline of the connections between the BELE model and the themes found in this literature review.

Table 3.
BELE and the Themes of this Review.

BELE Dimension		Description[1]	Themes from Review
1	Physical well-being and environmental comfort	Active learning, including significant emphasis on physical movement, sport, and outdoors activities used to promote learning and team work.	School safety School community Learning environments
2	Educational and cultural well-being	Active, creative, and participatory learning spaces, with high regard placed on individual difference. Integrated, multi-disciplinary learning in informal settings such as the school garden.	School community Learning environments
3	Socio-emotional well-being and the joy of learning	A friendly, joyful learning community, with a supportive environment where teachers and students collaborate to learn together.	School community Learning environments
4	Fantasy and innovation	The space to learn to become creative productive future citizens; open support and encouragement for innovation and fantasy in school work.	Learning environments Creativity

Notes. Adapted from Kangas (2010). [1] My paraphrase.

This school environment, designed by Finnish school children, would require a significant strategic change in staffing, operations, facilities, funding, etc., if it were to be implemented in U.S. public schools, but most importantly, would require a significant shift in thinking. In this model, the learner is truly at the center of the system. The caring school community provides a safe place for learning to occur. Learning happens using pedagogies and practices that Dewey (1938) and Kolb (1984) would instantly recognize as experiential learning, and creativity and creative practices are embedded in every aspect of the school.

Opportunity Statement

This review of the literature both supported and deepened my understanding of school climate, learning environments, and the teaching and learning of creativity in an elementary

school setting, culminating with Kangas' (2010) vision of the ideal BELE school. This new understanding of the domain reinforced my belief in the opportunity provided by this research. By enhancing the practice of school improvement with a specific focus on teaching and learning creativity, an opportunity emerges to better prepare young people to participate in the social life of the planet, thereby empowering the sustainability of human society. To achieve this, an action-oriented *theory of climate for creativity* is proposed, a measure of the theory constructed, and the intended uses and interpretations of its scores validated.

CHAPTER THREE:

MAPPING THE METHODOLOGICAL PATHWAY

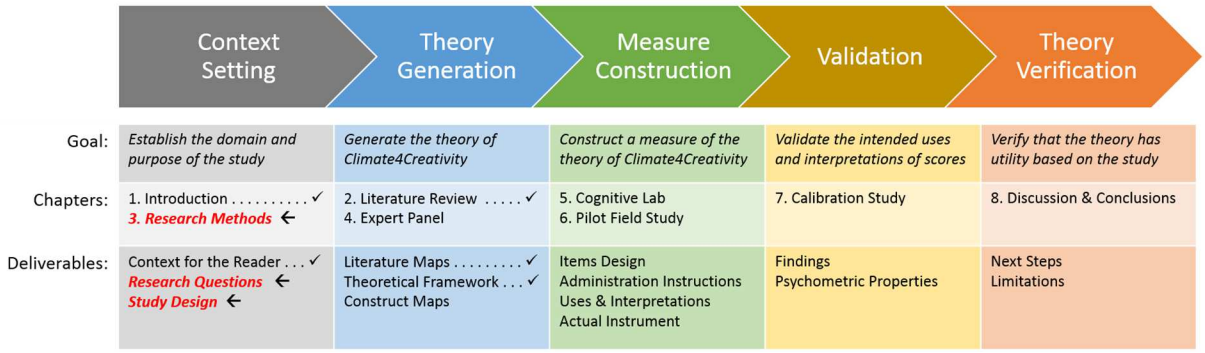


Figure 11. Dissertation Roadmap – Research Methods.

By enhancing the practice of school improvement with a specific focus on teaching and learning creativity, an opportunity emerges to better prepare young people to participate in the social life of the planet, thereby empowering the sustainability of human society. To achieve this, an action-oriented *theory of climate for creativity* is proposed, a measure of the theory constructed, and the intended uses and interpretations of its scores validated.

Research Design

In this research, measurement is used to define and enable change through an engaged process of examination, closely followed by a focusing process, designed to develop conjectures about what must change in order to make the lifeworlds of participants better. This focusing process leads to change action, where leaders and professionals can take positive steps, to integrate these new conjectures into the existent whole, thereby leading to new ways of being. This approach is iterative in nature, completion of a particular change action is followed by a

new measurement cycle, which in turn informs further focusing and action cycles. In order to guide this program, I developed the following series of four research questions:

RQ1: what are the core elements of healthy school climate & the environment for teaching and learning creativity in a public elementary school setting?

RQ2: how would we measure the degree to which elementary school students believe these core elements exist in their elementary school?

RQ3: what evidence of validity should be collected and what validation arguments should be applied to ensure credibility of such a measure?

RQ4: how might such a measure be employed to actually improve the environment for teaching and learning creativity in a public elementary school?

Exploratory Sequential and Concurrent Mixed Methods. This program of research used a sequential and concurrent mixed methods approach for the development, testing, and validation of the uses and interpretations of the scores from a pair of new psychometric instruments designed to measure student perspectives on school climate and the environment for teaching and learning creativity in a public elementary school. Figure 12 illustrates the overarching program design. Study I (expert panel) was completed in the summer of 2012. Study II & III (cognitive lab and pilot field study) were completed in the fall of 2012 and the spring of 2013. Study IV (calibration study) was completed in the summer and fall of 2013.

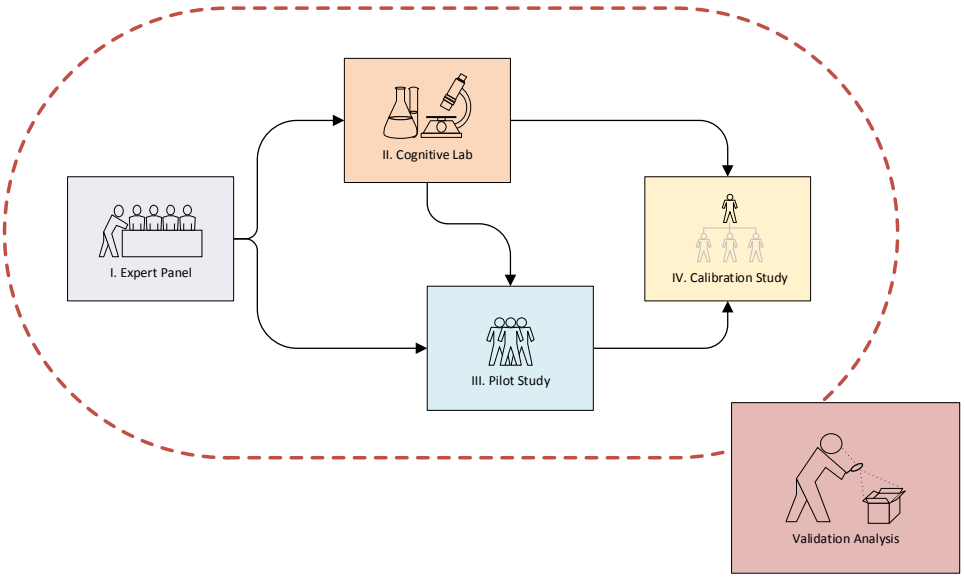


Figure 12. Four Study Mixed Methods Research Design.

Fixed Design with Emergent Internal Structure. This was a fixed research design, with each of the four studies having a distinct and important role in the overall program of research. The design was fixed in structure but simultaneously emergent in content, since the findings from each study purposefully informed the materials and procedures used in subsequent studies. The arrows in Figure 12 show the dependencies between the four studies.

Philosophically, study I represented a theory generation process, studies II and III involved construction and pretesting of a measurement instrument based on the theory, and study IV was a testing process for the measure and the theory. This research developed a new theory of school climate and the environment for teaching and learning creativity, *theory of Climate4Creativity*, and tested the theory empirically in the context of a large unified urban school district. Throughout the four studies, evidence was developed and collected for validation of the intended uses and interpretations of the scores from the Climate4Creativity Student

Perspectives Instruments for elementary (3-5) and middle school (6-8) grades (C4C/SPE and C4C/SPM).

Multiple Purposeful Worldviews. The research worldviews used to inform this design explicitly shifted from a constructivist perspective during theory development and refinement (studies I & II) to a largely postpositivist perspective during testing and calibration of the instrument (studies III & IV). Theory generation and instrument development generally used qualitative approaches to data collection and analysis, supported by some quantitative data, instrument testing generally used quantitative approaches to data collection and analysis. Altogether, this study embodied a pragmatic worldview of development, testing, and validation of two new psychometric instruments and to generation and verification of a new theory of action.

Matching Research Design to Research Questions. This research design aligned well with the exploratory, developmental nature of the research questions, progressively elaborating the theory of Climate4Creativity, the design of the measurement instrument, the application of the instrument in a calibration study, and the analysis of the results of that application. Table 4 illustrates the relationships between the research questions and studies within the program design.

Table 4.
Relationships between research questions and studies within the research design.

Research Questions	Study I: Expert Panel	Study II: Cognitive Lab	Study III: Pilot Study	Study IV: Calibration Study	Validation
RQ1: what are the core elements of healthy school climate & the environment for teaching and learning creativity in a public elementary school setting?	Define core elements			Test core elements	Validate core elements
RQ2: how would we measure the degree to which elementary school students believe these core elements exist in their elementary school?	Input to items design	Test / Refine items design	Test / Refine items design	Test items design	Validate items design
RQ3: what evidence of validity should be collected and what validation arguments should be applied to ensure credibility of such a measure?	Collect evidence of validity	Collect evidence of validity	Collect evidence of validity	Collect evidence of validity	Conduct validation analysis & develop validation argument
RQ4: how might such a measure be employed to actually improve the environment for teaching and learning creativity in a public elementary school?	Input to action methods		Input to action methods	Input to action methods	Define & document potential action methods

Reasons for Mixing Methods. Beginning this program it was essential to be open to emergent issues and insights, while controlling the time allowed for such emergence. In order to address these research questions, this design included highly emergent, generative phases (studies I & II), designed to generate a new theory of the climate for creativity and refine an items design for a measure of the newly generated theory. As standalone studies, these engagements served to develop emerging insights into the climate for creativity, but did not yield

particularly generalizable findings. These initial findings needed to be further tested in an attempt to create a more usable, generalizable tool for schools. This entailed use of a more deterministic perspective in Studies III and IV, empirically testing the instrument developed as an outcome from the theory generation process. Finally, a multiparadigm validation process built on the theory development and testing phases to document the evidence of validity collected during the preceding studies and to develop additional evidence where needed. This final validation analysis examined the intended and potential uses and interpretations of the scores yielded by the new instruments, with a focus on actionability of the measures.

Modern Validation Theory

Messick (1989) established a broadly defined *construct model* as the unifying framework for validity. In this model, construct validation would be used both to assess the justification for the use of an instrument (based on evidence or consequences) and to assess the outcome (interpretation or use of the scores) from the instrument. In other words, interpretations of scores should be validated *by reference to underlying constructs*, and, *uses of the scores should be supported by the connections between underlying constructs and intended uses*. This new unified definition of validity became known as modern validity theory,

> [validity is] . . . . an integrative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment. (Messick, 1989, p. 13)

Kane (2006) discussed a number of important implications emerging from this conception of validation, and from the philosophical evolution that led to it. First, modern validity theory emphasized the importance of making a clear and explicit statement of the proposed interpretations and actions based on measurement scores *in advance of validation*. The

focus had shifted from validation of the instrument as a way of assessing against a known criterion toward validation of the proposed interpretation and uses of the measures extracted from the instrument. For this reason, construct maps and clear statements of the intended interpretations and uses of the scores from the Climate4Creativity measures were created during the development process, and are presented in chapters 4 and 6 of this dissertation as part of the theory to be validated in this study.

Second, in order to ensure adequate validation, more comprehensive *programs of research* were required, rather than individual empirical studies. The underlying theory should be specified in terms of its constructs, measures should be developed for these constructs, and empirical evaluations of predictions based on the underlying theory should be conducted. This inevitably involved multiple phases of research within a carefully integrated research program. This program of research was carefully designed following this philosophical direction. In this case, four separate studies were integrated to achieve this result.

Third, because construct validation involved testing theories, there is a need to deliberately identify and assess alternate interpretations of test scores as part of the overall validation of an instrument (Cronbach, 1971; Kane, 2006).

Validation as Argument. When constructing an argument it is essential to take care that: 1) there is a clear thesis; 2) the arguments are coherent; 3) the arguments are supported by a variety of evidence; and, 4) the arguments are free from logical fallacies. By conceptualizing validation of an instrument in a similar way, the thesis is defined by the construct maps, items design, scoring guide, and measurement model created during instrument development; and is then validated through a set of arguments structured around evidence of the reasonableness of the intended interpretations and uses of the instrument. Under this conceptualization, there is no

checklist of validation criteria, there are, instead, a variety of ways to form an argument for the validity of the intended inferences and uses of a measure. Cronbach (1988) makes this clear,

> [the analysis]. . . . should make clear, and to the extent possible, persuasive the construction of reality and the value weightings implicit in a test and its application. (Cronbach, 1988)

Kane (2006) made the case for validation as classical argument, with the interpretive argument being similar to testing of a scientific theory (what he called "mini-theory," (Kane, 2006, p. 25). In his example validity argument for placement testing, Kane used an outline approach to define the argument, and then produced a written argument based on that outline. This approach allowed the instrument developer to clearly articulate the argument, and to assess the argument for clarity, internal consistency, and plausibility (particularly of inferences and assumptions). An argument structured in this way can be supported by call to pathos (credibility of experts in the literature review and in expert reviews of the instrument), logos (empirical data and logic derived from pilot testing and from calibration of the instrument), and by the ethos (credibility, style, and demeanor) of the instrument developer.

In this study, I used Mark Wilson's (2005) building blocks approach to measure construction, together with the modern validation theory approach embedded into professional practice in *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014).

Instrument Development as Validation

In modern validation theory, the instrument development process is viewed as a primary validation process. Developers use literature reviews and expert judgment to develop a theory of the domain under analysis, document this theory and specify a measurement procedure with an interpretive argument. The thesis is supported by a variety of arguments, each supported by expertise and empirical data gathered during the development process. In fact, the entire instrument development process can be viewed as a process of developing a thesis-argument structure designed to justify the interpretations and uses of an instrument. Kane (2006) summarizes this development stage as,

> much like the process of initial theory development and refinement in science, with the interpretive argument playing the role of a theory. The initial form of the theory is proposed. Any weaknesses identified in the theory are corrected, if possible, by changing some assumptions in the theory or by changing the scope of the theory (i.e., the range of cases to which it applies). If the evidence reveals inconsistencies that can't be resolved, the theory (or the interpretive argument) is rejected, but this is a last resort. The evidence produced during the development stage tends to be confirmationist; if a problem is identified, it is fixed, if it can be fixed. (p. 26)

Instrument development, therefore, *is* a mixed methods research process and requires a formal mixed methods research design. Wilson (2005) describes a structured four step approach to construction of a measure, an approach that purposefully uses a variety of approaches to both *gather evidence* and *be evidence* of validity for the intended uses and interpretations of the scores from the measure being developed. In this approach, both the *data* from early uses and the *process* used to develop the instrument are evidence of validity. An informal overview of measurement theories applicable to this work can be found in Appendix C.

Construct Mapping. Wilson's (2005) process begins with the development of *construct maps*. These maps are a structured specification of the theory or theories to be examined by the measure, providing a definition of the components of the measure, and the categorical scales for the scores which will ultimately be reported. The development of construct maps is a generally qualitative phase in instrument development, involving lit reviews, content panels, focus groups, interviews and other qualitative methods to develop a clear understanding of the theory or theories to be examined with the measurement instrument. The outcomes of this stage in the development process are construct maps, showing the components of a construct and describing the levels on those components. These construct maps are the theoretical embodiment of the line described by Wright & Stone (1979, p. 1), against which we intend to measure our respondents' perspectives. The construct mapping process in this work was implemented through an expert panel and initial literature review.

The Items Design. Next, the instrument developer creates an *items design* to manifest the theories embodied in the construct maps. Developers assert that responses to items are *caused* by the true level of attitude exhibited by each respondent. A variety of item types may be used to manifest the constructs found in the construct maps. Development of the items design is a mindful, creative act purposefully developing and refining items based on the constructs being examined by the measure under construction. This is a generally qualitative phase, informed by the literature, by empirical research on the efficacy of specific item types in specific circumstances, and bounded by the requirements of the construct maps developed in the first stage. This design work is generally supported by a small scale quantitative data collection, examining the initial items design in actual practice through a pilot study, and using a focused psychometric analysis to identify items for further investigation and refinement. In most cases,

the construct maps are adjusted during this phase, in an iterative way. This iterative refinement is a normal part of the creative process used to develop an instrument.

The Outcomes Space. Once an items design has been completed, an *outcomes space* analysis is conducted. In this phase, a scoring guide is developed, mapping potential responses to items in the items design to numerical *raw scores* for the instrument. These raw scores are not the inference themselves, but begin the process of developing a model of inference for the measure. This outcomes space is a pivot point in the development of an instrument: it begins to map the qualitative theory-manifestation (construct maps) to what will ultimately be a quantitative measure (construct measures) – it prepares the instrument design for its intended inference. Depending on the item types used in the items design, this may be a simple numerical mapping process or may require development of detailed scoring rubrics and procedures, from which trained raters can consistently score the responses received to the items in the measure (Wilson, 2005).

The Measurement Model. Finally, having developed a scoring guide for the measure, the interpretation structure or *measurement model* is created, to map scores from item responses to actual locations on the construct maps. Figure 13 elucidates the relationship between the construct maps as *theory specification*, causing the items responses (*theory manifestation*), being scored via the *outcomes space*, and allowing *inference of location* on the construct maps through the measurement model. The response to a valid items design will not only be *caused* by the construct, but will allow the researcher to *infer* evidence for the construct itself, based on the scores from the measure, connected back to the construct maps.

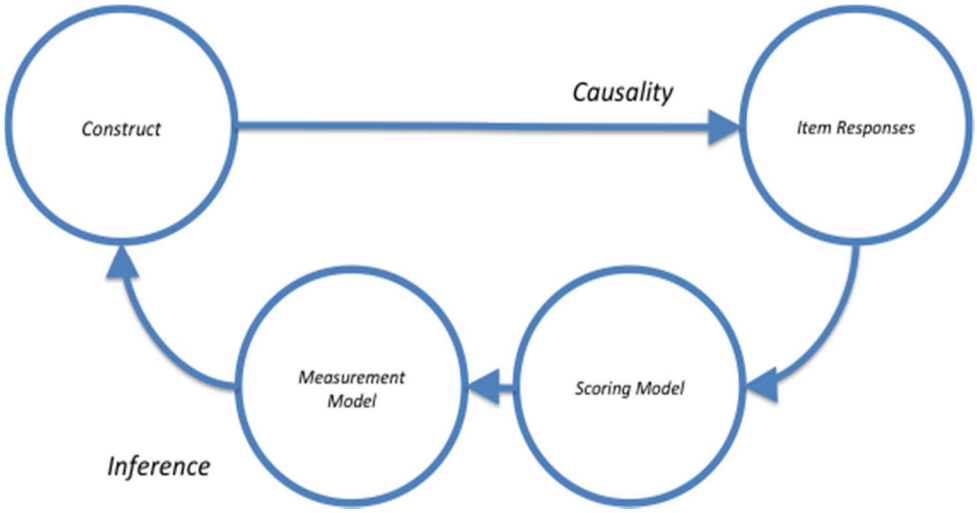


Figure 13. Wilson's (2005) Four Building Blocks approach to Instrument Development.

Note. Adapted from Wilson (2005, p17).

Construct maps for the theory of Climate4Creativity can be found in chapter 4, along with a narrative description of the theory. Chapter 6 includes information about the completed items design (the full specification can be found in Appendix E). A statement of intended uses and interpretations of the scores yielded from the instruments can be found in chapter 6.

Evaluative Study as Validation

The validation of the intended interpretations and uses of the scores yielded by an instrument begins in the instrument development process and continues in evaluative studies conducted following the first significant use of the instrument. The process of validation continues for the full useful life of the instrument, each time it is used, its use should be validated to ensure it continues to be an effective tool for examination of the issue at hand (Messick, 1989). In this context, a framework for validation has utility to ensure that the evaluator examines the most important aspects of the instrument's interpretation and use in the major study. For the purposes of this dissertation, the 2014 edition of *Standards* (American

Educational Research Association et al., 2014) was used, with particular focus on standards 1.1 through 1.25. Not every standard was applicable to this study, but the framework provided by these standards was used as a guide and organizing structure for development of the interpretative validation argument. See Appendix K for a description of standards applicable to this study.

Standards (American Educational Research Association et al., 2014) describe five sources of validity evidence for a study. *Evidence based on test content* provides an analysis of the interpretations and uses of the scores yielded by a measure, based on the processes used to develop and describe the content being measured, and the relationships between the items in the measure and the content being measured. *Evidence based on response processes* examines the cognitive processes of the test takers, using theoretical and empirical evidence of the fit between the constructs being measured and the actual performance of the respondents. *Evidence based on internal structure* examines the internal consistency of the instrument, and evaluates the degree to which the items design fits the model of the constructs being measured. *Evidence based on relations to other variables* assesses the degree to which major components of the instrument match with other external measures and variables, focused on the same or similar topics. *Evidence for validity and consequences of testing* examines the consequences of testing, with particular focus on evidence of the soundness of proposed interpretations and their intended uses (American Educational Research Association et al., 2014, pp. 11-20). These five sources of evidence should be integrated into a sound validity argument, a "coherent account of the degree to with existing evidence and theory support the intended interpretation of test scores for specific uses" (p. 21). The validation argument developed in this dissertation uses these standards as its guiding principle.

CHAPTER FOUR:

GATHERING EXPERT PERSPECTIVES

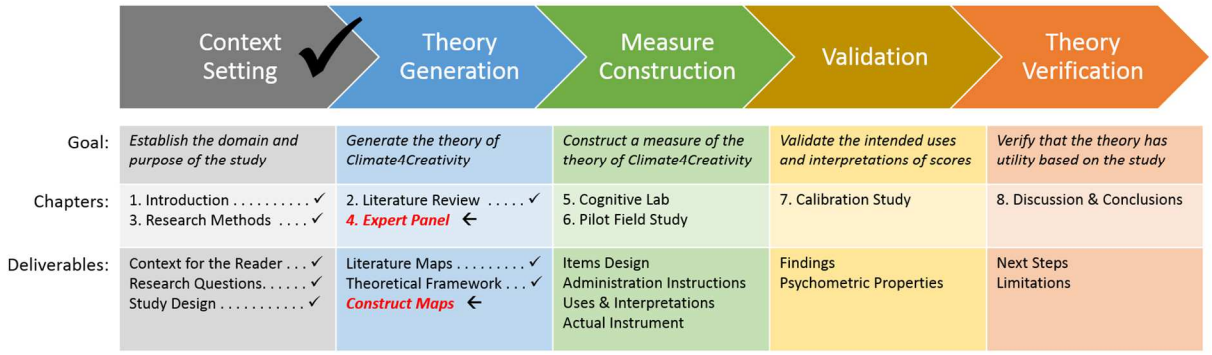


Figure 14. Dissertation Roadmap – Expert Panel.

By enhancing the practice of school improvement with a specific focus on teaching and learning creativity, an opportunity emerges to better prepare young people to participate in the social life of the planet, thereby empowering the sustainability of human society. To achieve this, an action-oriented *theory of climate for creativity* is proposed, a measure of the theory constructed, and the intended uses and interpretations of its scores validated.

Study I: Expert Panel

This study was the first phase of a mixed methods research program designed to explore school climate and the environment for teaching and learning creativity in an elementary school setting. The study was designed to provide expert input to the following research questions:

RQ1: what are the core elements of healthy school climate & the environment for teaching and learning creativity in a public elementary school setting?

RQ2: how would we measure the degree to which elementary school students believe these core elements exist in their elementary school?

RQ3: what evidence of validity should be collected and what validation arguments should be applied to ensure credibility of such a measure?

RQ4: how might such a measure be employed to actually improve school climate & the environment for teaching and learning creativity in a public elementary school?

Introduction. The process used in this study was to work with a collection of experts to gather input and feedback to refine and improve a set of initial construct definitions and an initial items design for the Climate4Creativity Student Perspectives instrument and to begin to develop and document evidence of validity of that instrument for its intended uses and interpretations. The expert panel first completed the existing Comer School Climate Survey as a reference instrument (Yale School of Medicine, 2009), and then assessed the draft constructs and candidate items for the new instrument using a paper survey and a facilitated group discussion process.

Following this study, a series of additional studies were conducted, as defined in the research design, to develop and test the instrument for use by the Bridgeport, CT school system to assess school climate and the environment for teaching and learning creativity in Bridgeport schools. This was part of a partnership with the Bridgeport schools, designed to enable the schools to meet the requirements of Connecticut state regulation on school climate (An Act Concerning the Strengthening of School Bullying Laws of 2011) and to gather and analyze data that would be used by the school district to improve the climate for learning in the schools.

Participants. Seventeen expert participants took part in the expert panel. Of these seventeen participants, two were current or former school principals, nine were certified teachers, seven were creative professionals, eight had experience as school support professionals,

three had quantitative research backgrounds, four were Bridgeport school community members, and ten were parents of school age children. There were eight men and nine women on the panel, with educational backgrounds ranging from two individuals with some college experience through six individuals with terminal degrees (two MFAs, one MSW, two Ed.Ds, and one Ph.D). Two individuals had learning disabilities and one was physically disabled. One openly gay man participated. Ages ranged from 27 to 64 years of age, with an average age of 47. This was a diverse and engaged panel of content and methods experts. About half the panel had never met any of the other panelists.

Procedures and Materials. This expert panel took place in the Cambridge Commons area at Sacred Heart University in Trumbull, CT on July 10, 2012. Following IRB approval of the study, a contact list of approximately 80 candidate experts living and working in Connecticut was developed using personal contacts, and expanding the search by asking contacts and advisors to recommend additional participants. Three preparation emails were sent to the list of candidate participants: an initial contact email asking candidate experts to express interest in the event, an invitation email asking interested candidates to agree to attend the panel event, and finally a preparation and pre-read email about a week prior to the panel event. Informed legal consent was obtained from all participants, either in advance via an electronic consent form, or on site at the event. A detailed demographic form was also obtained from all participants. Information about the study was provided in advance of the panel event as part of the pre-read materials, and additional information was available on site. In order to protect the confidentiality of participants, no attendee list was published, each participant was allocated a number on arrival at the event, and only first names were used in facilitation of expert activities.

Upon arrival at the event, each participant was given a large envelope containing her study materials. Materials were returned to these envelopes and sealed by the participant at the conclusion of the study, and were then stored in a locked cabinet in my office on the campus of Albertus Magnus College in New Haven, CT. Participants were not paid for their involvement, but were provided with refreshments at the event. Participants were given information about the study in a brief presentation and there were opportunities to ask questions and discuss the procedures used in the study. Participants completed a demographic information form and signed an individual informed consent form before beginning the work of the study.

Data Collection. Having completed the informed consent process, participants completed a paper version of an existing reference instrument – the Comer Elementary and Middle School Climate Survey (Yale School of Medicine, 2009). This reference instrument served two major purposes in the study. First, it provided a mechanism to engage participants in thinking about school climate and healthy learning environments; and, second, it provided a small scale parallel test using a different but comparable instrument for three of the four constructs being examined.

Following completion of the Comer instrument, participants completed four *item response assessment surveys (IRAS)* instruments, each examining the initial items design for one of the four constructs in the new instrument. The IRAS instruments were, in essence, metainstruments – instruments designed to collect qualitative and quantitative data about another instrument. Respondents were asked to complete the four IRAS instruments in four different sequences: ABCD, BCDA, CDAB, DABC; so as to ensure that all four constructs had at least some responses, even if respondents ran out of time to complete all their individually assigned instruments.

Respondents first answered questions on each proposed item, addressing the *difficulty* of answering each item using the response options provided. Specifically, respondents were asked to create a written *think aloud* (Willis, 2005) about how difficult it was to translate their true feelings about an item using the options provided, and to score the difficulty of each item on a 1-7 scale, with 1 meaning *really easy to answer* and 7 meaning *really difficult to answer*. Second, respondents were asked to review how *relevant* they felt the item was to the construct and dimension being measured, once again, respondents were asked to think aloud (Willis, 2005) about the relevancy of the item, and this relevancy was scored on a 1-7 scale with 1 meaning *not relevant*, and 7 meaning *perfectly relevant*. Respondents were given free form space to make any additional comments they had on the difficulty and relevancy of the question. The verbal instructions given to participants encouraged them to fully express their perspective as openly and honestly as they were able, and to use additional paper as necessary.

Having analyzed and provided input to individual items, respondents were asked to assess the overall items design, addressing the *representativeness* of the items to the construct being measured. Panelists scored the representativeness on a 1-7 scale with 1 meaning *items do not represent the construct* and 7 meaning *items completely represent the construct*. Participants were then asked to edit the item set by amending, deleting, reordering, or adding items to the construct. Once again, participants were asked to provide free form feedback on any aspect of the overall items design, and to use additional paper as necessary.

Having completed their four item response assessment surveys, participants were engaged in a facilitated discussion focused on five key questions. Participants were first asked to write answers to the questions on sticky notes, in order to engage the more introverted members of the panel in complete thinking, before a verbal discussion began. As each set of sticky notes

was placed on the board by participants, a discussion took place to probe and explore panelists' perspectives on the topics at hand, thereby enabling and encouraging panelists to interact and build new meaning from one another's comments.

The following five questions were used to guide the facilitated discussion:

1. *What do you like about this instrument?*
2. *What things would you like to see changed about this instrument?*
3. *What challenges do you foresee to implementation of this instrument?*
4. *What value do you see to Bridgeport stakeholders from the use of this instrument?*
5. *What is your gut reaction to what you have seen this evening?*

Finally, prior to leaving the event, participants were asked to complete a feedback form designed to assess their experience at the research event itself, and to invite continued participation on future research work. Envelopes were collected, checked for completeness, resealed, and placed in a portable lockbox which was transported to my office on campus at Albertus Magnus College in New Haven, CT, where envelopes were stored in a locked cabinet for analysis.

Data Analysis. The 17 sets of item response assessment surveys were organized by construct and individually reviewed. Scores for difficulty and relevancy of each item were entered into an Excel spreadsheet (Microsoft, 2013a). Items of high relevancy and low difficulty were marked for continuation, items of high relevancy and high difficulty were carefully reviewed, using the free form comments from respondents to understand the root causes of the difficulty and relevancy scores. Medium and low relevancy items were reviewed to assess if they could be improved or should be discarded. Several follow-up email questions were sent to respondents to clarify comments on their surveys. Missing items noted on forms were added into

the spreadsheet as potential new items. Structural comments were noted in a separate worksheet within the results spreadsheet, and later thematically coded to identify important areas of improvement. Over 400 individual points of feedback were extracted from the item response assessment surveys, and a number of major themes emerged through iterative analysis of the comments. Notes from the facilitated discussion were captured in an Excel (Microsoft, 2013a) spreadsheet and analyzed to group responses into thematic responses to each question.

Item Response Assessment Survey Results. In addition to over 400 individual points of feedback or clarification, six major themes emerged from the open-ended comment responses on the item response assessment surveys, described below.

Theme 1: Multiple Classes in Middle & High School. Participants noted that many of the items included questions about the quality of teaching, interest or effort in a classroom. These questions were difficult to answer in the abstract, without specifying for the respondent *which specific teacher or classroom* they referred to. Suggestions included providing a table of subject areas and asking students to answer for all subjects, having students pick their favorite subject and answer for just that one, or randomly assigning subjects to respondents to provide roughly equal samples for each subject area.

Theme 2: Wording with Target Audience in Mind. Multiple participants noted challenging or academic wording in some items. One example is the use of the words *task* and *skills* in dimensions D2 and D3. In each case, a number of alternate wordings were offered by participants. Several participants suggested having items reviewed by a group of comparable age students to the target audience to ensure that wording meant the same thing to elementary school respondents as it did to adult education experts.

Theme 3: Options Not Mutually Exclusive. In several cases, participants noted option-sets that were not mutually exclusive, leading to difficult choices. In some cases respondents felt they could select multiple options, in others they felt that the differences between two or more options were not significant enough to differentiate. Suggestions included using a checklist of options and tightening the wording to ensure mutual exclusivity. Once again, several respondents suggested testing the options with target age respondents to ensure they could appropriately differentiate between options.

Theme 4: Question-sets for Students with Disabilities. Particularly in the creativity items, several participants commented on the use of the words *distraction* and *focus* as partial criteria for an environment for creativity. Several respondents felt that since many students have ADHD and other attention related disabilities, these questions might not yield valid results. Some students simply may never be able to focus due to their disability, even in an environment where creativity is highly supported.

Theme 5: Creativity Question-set. Most participants were disappointed with the depth of the creativity items, commenting that they were *light* or that they did not seem to cover the entire space of creative environment. Three respondents called for a complete redesign of the creativity construct, providing specific input to the structure of the construct and items design.

Theme 6: Anonymity vs Identification. Several participants commented that students may not tell the truth on the instrument, for fear of repercussions for their statements. Particularly on the bullying and teacher quality items, where students might feel that a negative response might lead to retaliation from teachers or peers. A second, related discussion on this theme centered around how to identify specific students in order to address any student reports of bullying *happening now*, versus the need to maintain confidentiality in order to encourage

students to reveal their true feelings. A concern that this instrument could be inappropriately used by school administrators in teacher evaluation was raised by one expert, and echoed by several others.

Facilitated Discussion Results. In addition to over 120 individual points of feedback or clarification, a number of themes emerged from the oral discussion of each question, as follows.

Question 1: What do you like about this instrument? A total of twenty-nine sticky notes were posted as part of this discussion, and a 10 minute conversation was facilitated among respondents on this topic. Eight respondents particularly liked the idea of integrating the environment for creativity into the constructs of school climate, these respondents felt that the connection between the items and the constructs was clear, and that adding creativity was a positive move in understanding school climate. Seven respondents felt that the varied format of the instrument, with both multiple choice and open-ended responses was positive. Six respondents felt that the broad scope of the survey made the instrument more comprehensive than other climate surveys they had seen. Five respondents felt that the instrument would make students think about their learning environment and might encourage them to change their behavior. Two respondents felt that the survey would provide excellent feedback for school administrators in managing the climate in their schools. Oral discussion centered on the positive impact the instrument could have if used to help schools understand the student perspective and improve school climate, thereby improving educational outcomes for their students. Several panelists commented on the degree that students are more aware of their learning environment than adults give them credit for.

Question 2: What would you like to change about the instrument? A total of thirty-seven sticky notes were posted for this question, including four giving feedback on the item

response assessment survey itself, relating to the coding of dependent item logic at the start of each question, a consequence of the paper survey format. Eight participants suggested adding broader perspectives than the student perspective, including teacher, parent, and school staff perspectives. Seven participants wanted to simplify wording and delineate items with more clarity. Seven participants suggested reducing the number of available options or allowing respondents to pick more than one option from lists of answer choices. Seven respondents suggested missing areas of focus in the items design. Three participants wanted more depth and quantity of items related to creativity. One participant suggested creating a *report bullying now* option, allowing respondents to jump to a page where they could file a bullying report and have it immediately emailed to the anti-bullying coordinator for their school. The consensus on the panel was that the instrument should remain strictly anonymous, and focused on its intended use, to improve the environment overall, rather than identify scores for specific students.

Question 3: What challenges to implementation do you foresee? A total of fifteen sticky notes were posted for this question, and this was the most vibrant discussion amongst expert panelists. Eight respondents were concerned about getting buy-in from schools and from the district and finding time to complete the instrument during a busy school day, especially if there are no action plans in place for what to do with the results of the measure. Three respondents felt that improving the wording of items would be a major challenge. Two panelists felt that budgetary constraints would probably squash this initiative. Two panelists felt that it would be difficult to make the case for creativity in a school system like Bridgeport, one commenting, *how can we possibly teach creativity in a school system where students can barely read?*

Question 4: What value does this instrument hold for Bridgeport stakeholders? A total of twenty-six sticky notes were posted on this question and a vibrant discussion ensued.

Comments were organized into four groups of stakeholders – Students, Parents, Teachers, and Other Stakeholders. Two panelists felt that students would feel they were given a voice in their school, this point was emphasized by several experts in oral discussion. Two panelists felt that students would be engaged in holding others accountable, and two felt that students would see improved outcomes in terms of school climate at their schools as a result of completing this survey. One panelist felt that the survey would be fun for students to complete. Three panelists felt that parents would benefit from having happier kids, and would better know what was going on in their schools. Five respondents felt teachers would benefit from understanding the relationships between creativity and academic outcomes. Two respondents felt that this survey might help to enable teachers to be creative and have fun at school. Three respondents felt that this survey would provide data to improve teacher prep programs. Four panelists felt that the data would have great value for district administrators, and three felt that the surveys would help Bridgeport schools become a model for turnaround amongst challenged school systems.

Question 5: What is your gut reaction to this project? A total of sixteen sticky notes were posted related to this question. Nine of these were highly supportive of the project, specifically offering support for ongoing work, and using phrases such as, *excellent start*, *keep going*, and *go for it!* Three respondents felt that a clearer definition of creativity was needed to eliminate confusion around creativity in the arts versus a broader conception of creative thinking. Two respondents felt that the project would likely be crushed by budget priorities, and two felt that more feedback/work was needed before the instrument was ready to go. The overall conversation was highly positive and supportive of moving forward with the instrument, provided appropriate care was taken to ensure the validity of the instrument.

Discussion. The primary purpose of this study was to develop and refine a set of initial constructs for the theory of Climate4Creativity, which would be used to design the Climate4Creativity Student Perspectives (C4C/SP) instruments. In order to achieve this, the study defined, refined, and selected items to be included in the instrument. The item response assessment survey was seen by participants as a highly effective tool to achieve both purposes, although several participants felt that the process of defining, refining, and selecting items to be included was stronger than the construct development process. In addition, thematic analysis of the free form comments on the survey emerged as an important feature of the study, providing the impetus for several changes to the constructs and to the structure of the instrument.

Analysis of *Theme 1: Multiple classes in middle & high school* led to an important structural change to the instrument. In the initial version of the instrument, the question-sets did not specify exactly which classrooms were being referenced when discussing the learning environment or teacher efficacy. This made it difficult for respondents to consistently connect their responses to their actual experience in classrooms, particularly in the higher elementary grades (grades 6-8), where students take a variety of subjects, often with different teachers in different classrooms. After analysis, I determined to resolve this by adding two additional items at the start of the grade 6-8 survey question, asking students to select their favorite and least favorite subjects from a list of all available subjects at the school. The learning environment and creativity construct questions would then be asked twice, using the favorite and least favorite subjects as drop in text in the questions. This way, student respondents would be asked to, think about your *math* classroom… prior to a bank of questions about learning and creativity in that specific classroom. This approach would provide clarity of the learning environment and creativity items for that particular student, and could be potentially combined into a single

measure, perhaps using an average of the two scores (the average of the favorite and least favorite classroom). The grade 3-5 item would be left as a single classroom item, since these students are generally instructed by a single home room teacher.

The issue raised in *Theme 2: Wording with target audience in mind* led to an in depth review of the language used in each item. Some changes were immediately implemented, to make questions more consistent and to use simpler language. These changes would be further refined in the cognitive lab study to follow.

Theme 3: Options not mutually exclusive, led to a careful analysis and redesign of the answer choices on several items across the survey, all multichoice answer choices were converted to 4-point or 5-point scales on a carefully drawn hierarchy of choices. Wright maps (Wilson, 2005) were used to document these choices for each item. Following development of these Wright maps, a careful review was completed to ensure that options were mutually exclusive and collectively exhaustive in every case.

Taken together, *Theme 4: Question-sets for students with disabilities* and *Theme 5: Creativity question-set* led to a almost complete redesign of the items in the creativity construct, building a deeper and more complete version, and simultaneously eliminating words which might be difficult for a student with an attention related disability to answer.

Concerns raised in *Theme 6: Anonymity vs identification* led to a clear statement on anonymity of responses. No personally identifiable information would be captured, and respondents would be clearly informed by classroom moderators that individual responses will not be shared with their school. This would become a fidelity criteria in all technical documentation. In reporting, this added an ethical challenge, since, if an individual student were to report that she was *being bullied now*, we would be *unable* to identify who had reported being

bullied and could not provide this information to school antibullying coordinators. Based on an idea discussed in the facilitated discussion, a hyperlink was added to the survey at the bottom of the bullying questions page, to inform the student of their right to support from their school, and to let the respondent know the name and phone# of the antibullying coordinator at their school. While this does not fully resolve the ethical dilemma, it does provide some important information to the individual student, and it does make it clear that confidentiality will be rigorously maintained.

The question of use of this instrument in teacher evaluation warranted a clear response. This instrument will not be designed for teacher evaluation, and will not be evaluated for use in teacher evaluation. Responses will not be collected at the level of identified teachers, but at the grade and subject level, providing only a limited and indirect connection between student responses and teacher performance. The use of this instrument in teacher or principal evaluation will not be validated as part of this dissertation study.

A number of interesting responses from the facilitated discussion led to changes to the items design, particularly in the items related to the creativity construct. The suggestion to broaden perspectives to include teacher, parent, and school staff input was taken on board, and is currently planned as a post-doctoral project, following completion of this program of research. The suggestion raised by several respondents, that an evidence based action plan for each key response be put in place in advance of collection of data led to the creation of a district intervention action planning team, led by the district PBIS team leader, Mayra Perez. The purpose of this district team was to create a document mapping groups of items included in the instrument to proposed interventions, based on the team's research into evidence based practices.

This document provided recommendations to district and school leaders, as input to their individual school action plans.

Conclusions. Overall, this study led to several important clarifications and adjustments to the draft theory of Climate4Creativity (manifested through construct mapping), and to a significant redesign of the items in the initial draft Climate4Creativity Student Perspectives instrument. Version 1.0 was updated to version 1.1, with version 1.1e designed for elementary grades 3-5, and 1.1m designed for middle school grades 6-8. The updated items design can be seen in Appendix F. Of significant note were four critical items design decisions, based on the feedback from the expert panel.

Decision 1: Split the instrument by age into an elementary and middle school instrument version, keyed around the point when students begin to take classes with multiple teachers in multiple classroom environments. This would generally mean grades 3-5 take the elementary school version and grades 6-8 take the middle school version. The core questions would be the same, but some supplementary questions would be asked to older students but not younger students. In addition, younger students would only be asked learning environment and classroom level questions once, whereas older students would be asked these questions twice: once for their favorite subject and once for their least favorite subject. This led to a core elementary school items design with a few additional questions which were only asked to older, middle school aged, students.

Decision 2: Removal of free form items in favor of more comprehensive use of multiple choice items. Based on the discussion, experts generally felt that, provided items were adjusted to ensure mutual exclusivity and completeness of each set of answer choices, it would be beneficial to use a multiple choice approach with students, particularly in today's schools, where

multiple choice tests and questions were quite typical for students. Experts suggested that some freeform options could be provided for older students, perhaps in high school, if the instrument were to be expanded into that age range.

Decision 3: Number of multichoice options. A debate took place at the panel, about the preferred number of answer choices for multichoice items. This was left as an open question at the end of the panel, since there was not consensus amongst the experts on the panel. Most felt that between 4 and 7 options were appropriate, with less options for younger students, but that was the extent of the consensus. The issue of whether to provide a natural 5- or 7- option scale or a forced choice 4- or 6- option scale was raised but not resolved at the panel. Since Wright maps were used extensively in development of the construct definitions for this instrument, and these had naturally been generated with four or five point scales, I decided to test a four point scale with the younger students (elementary grades 3-5) and a five point scale with the middle school students (grades 6-8), with the aim of ultimately selecting the same base scale for both instruments. Version 1.1 of the instrument therefore uses a 4-point scale for its elementary design and a 5-point scale for its middle school design, with, essentially, the same items with slightly different answer choices.

Decision 4: Integration of learning environment and creativity questions into a more seamless flow of questions. Experts felt that rather than asking all the safety, then community, then learning, then creativity questions, it would reduce survey fatigue if the questions were carefully sequenced to provide interest and movement for the respondent. Questions could then be extracted and applied to the constructs after the fact. This decision simplified the design, removed the need for some of the more complicated if-then-else logic, and, ultimately, shortened the survey.

Limitations. Due to the overwhelming feedback received on the items design for the creativity construct, a complete redesign of several key components was conducted. A second expert review of this new items design would have added a level of confidence to the next steps. The new design was, however, triggered by expert feedback, and took account of extensive information provided by the experts, as well as a deeper review of creativity literature following this initial event. A more extensive discussion of the constructs prior to completion of the item response assessment survey process might have made it easier for the experts to comment on the constructs, in addition to the study's focus on the items and answer choices. This did not stop this particular group of experts from expressing their opinions about the constructs throughout the panel event – however, a more explicit review of the constructs might have had some additional benefits as a primary activity. Finally, the reference survey completion was a lost opportunity. Respondents provided extensive feedback on the Comer reference instruments, particularly on perceived failings and poor quality of items in the reference instruments. At this point, these were taken as a series of warnings about what *not* to do in the instrument under development. These could, perhaps, have been used to help enable a deeper discussion of the new instruments.

Construct Mapping

Construct mapping is a systematic approach to defining the theoretical structure underlying a measurement instrument, encapsulating the theory being measured by the instrument into a series of maps which describe the levels of a measurement scale using qualitative descriptions of each level. These levels are later translated into quantitative measures by use of a measurement model. The theory of Climate4Creativity, as defined following the expert panel study, included four constructs – School is a Safe Place; School is a Caring

Community; School is a Place for Learning; and, School is a Place for Creativity. Each core construct was explained as an ordinal scale, from a low level to a high level, and a summary measure of the Climate4Creativity was extracted by combination of the core construct measures. See Figure 15 for a structural view of the constructs embedded in the theory of Climate4Creativity.

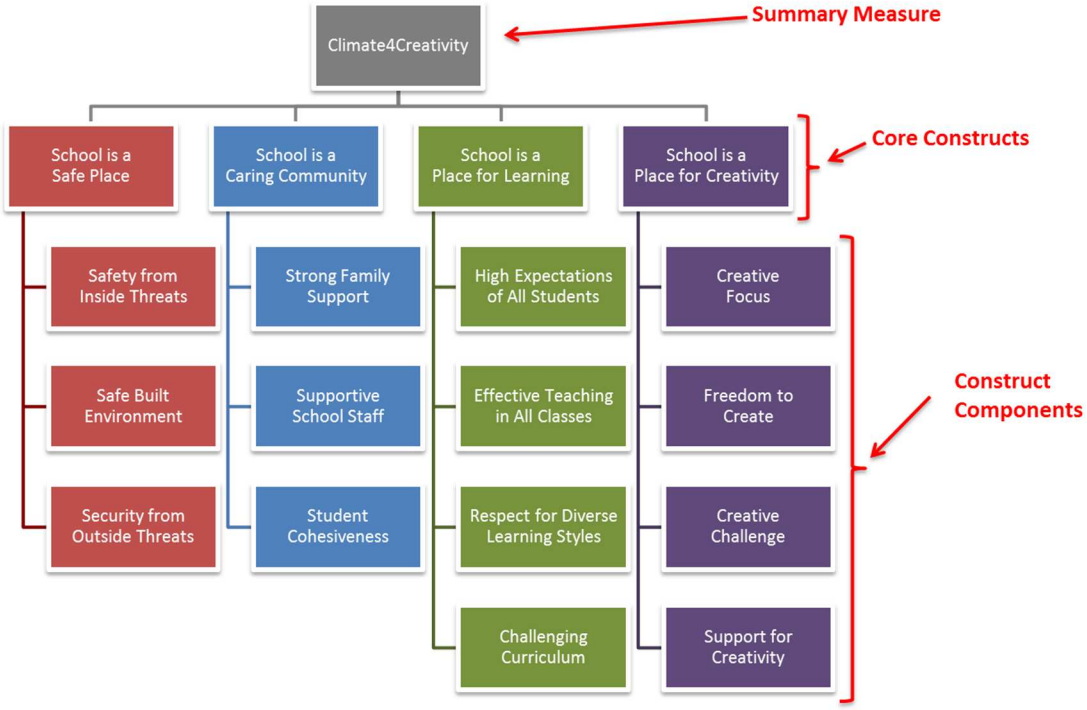


Figure 15. The Climate4Creativity Construct Structure.

School is a Safe Place. The first core construct in the Theory of Climate4Creativity was defined as School is a Safe Place. This core construct had three components, *Safety from Inside Threats*, *Safe Built Environment*, and *Security from Outside Threats*, and was embodied by a scale ranging from a low level of safety to a high level of safety. The School is a Safe Place construct map is shown in Table 5.

Table 5.
Construct map for the School is a Safe Place construct.

Level of Construct	Safety from Inside Threats	Safe Built Environment	Security from Outside Threats
Very Safe	There is no bullying in my school	I feel completely safe everywhere on school grounds	I completely trust my school to keep me secure from outsiders
	School is always effective at protecting me	I feel completely safe all day long, regardless of school activity	
Safe	I rarely experience, and rarely see or hear about bullying at school	I feel safe everywhere on school grounds	I trust that my school will protect me from outsiders
	School is usually effective at protecting me	I feel safe all day long, regardless of school activity	
Mostly Safe	I sometimes experience and sometimes see or hear about bullying at school	I feel safe most places on school grounds	I feel that my school mostly keeps me protected from outsiders
	School is mostly effective at protecting me	I feel safe most of the time during the school day	
Somewhat Unsafe	I often experience, see or hear about bullying at school	I feel somewhat unsafe in some places on school grounds	I feel that my school sometimes does not keep me protected from outsiders
	School is not very effective at protecting me	I feel somewhat unsafe at different times during the school day	
Very Unsafe	I often experience, and constantly see and hear about bullying at school	I feel unsafe in a lot of place on school grounds	I do not trust my school to protect me from outsiders
	School is not effective at protecting me	I feel unsafe a lot of the time during the school day	

School is a Caring Community. The second core construct in the Theory of Climate4Creativity was postulated as School is a Caring Community. This core construct had three components, *Strong Family Support*, *Supportive School Staff*, and *Student Cohesiveness*, and was embodied by a scale ranging from a low level of caring community to a high level of caring community. The School is a Caring Community construct map is shown in Table 6.

Table 6.
Construct map for the School is a Caring Community construct.

Level of Construct	Strong Family Support	Supportive School Staff	Student Cohesiveness
Love My School	My family are constantly involved in school	My teachers and other adults at school always help, support, and encourage me to succeed	Other students always help, support, and encourage me to succeed
	My family push me to succeed at school	My teachers always treat me with kindness and respect	Students are always kind and respectful to each other
Like My School	My family are very involved in school	My teachers and other adults at school usually help, support, and encourage me to succeed	Other students usually help, support, and encourage me to succeed
	My family help me to succeed at school	My teachers usually treat me with kindness and respect	Students are usually kind and respectful to each other
My School is OK	My family are involved in school	My teachers and other adults at school sometimes help, support, and encourage me to succeed	Some other students help, support, and encourage me to succeed
	My family help me to succeed at school	My teachers often treat me with kindness and respect	Students are sometimes kind and respectful to each other
Dislike My School	My family are not very involved in school	My teachers and other adults at school don't support me	Other students rarely help, support, and encourage me to succeed
	My family don't help me much with school work	My teachers rarely treat me with kindness and respect	Students are rarely kind and respectful to each other
Hate My School	My family are not involved in school	My teachers and other adults at school don't care if I succeed	Other students don't care if I succeed
	My family do not support me at school	My teachers never treat me with kindness and respect	Students are mean and disrespectful to each other

School is a Place for Learning. The third core construct in the Theory of Climate4Creativity was defined as School is a Place for Learning. This core construct had four components, *High Expectations of All Students*, *Effective Teaching in All Classes*, *Respect for Diverse Learning Styles*, and *Challenging Curriculum*, and was embodied by a scale ranging from a low level of general learning environment to a high level of learning environment. The School is a Place for Learning construct map is shown in Table 7.

Table 7.
Construct map for the School is a Place for Learning construct.

Level of Construct	High Expectations	Effective Teaching	Diverse Learning Styles	Challenging Curriculum
Learning More Than Expected	My teacher expects excellent work and inspires me to do better	My teacher always gives clear direction	I learn in many different ways at school	My school work is always difficult
	I expect to earn excellent grades	I always finish my work in class	I often get to work in teams with other students	
Learning as Directed	My teacher expects good work and inspires me to do better	My teacher usually gives clear direction	I learn in different ways at school	My school work is usually difficult
	I expect to earn good grades	I usually finish my work in class	I get to work in teams with other students	
Learning Most of Expected	My teacher expects OK work	My teacher mostly gives clear direction	I learn in a few different ways at school	My school work is about right – not too difficult, not too easy
	I expect to earn OK grades	I often finish my work in class	I sometimes get to work in teams with other students	
Not Learning Much	My teacher doesn't expect much from me	My teacher is sometimes unclear	We usually learn in the same way at school	My school work is usually pretty easy
	I expect to earn below average grades	I rarely finish my work in class	I rarely get to work in teams with other students	
Learning Nothing	My teacher expects me to fail	My teacher never gives clear direction	Every class is the same	My school work is always very easy
	I expect to earn failing grades	I rarely finish my work in class	I never get to work in teams with other students	

School is a Place for Creativity. The final core construct in the Theory of Climate4Creativity was defined as School is a Place for Creativity. This core construct had four components, *Creative Focus*, *Freedom to Create*, *Creative Challenge*, and *Support for Creativity*, and was embodied by a scale ranging from a low level of creative environment to a high level of creative environment. The School is a Place for Creativity construct map is shown in Table 8.

Table 8.
Construct Map for the School is a Place for Creativity Construct.

Level of Construct	Creative Focus	Freedom to Create	Creative Challenge	Support for Creativity
Creativity is Required	The work I do in class is always fun	I always rethink or reframe class problems	I always need to think deeply about class problems	My teacher requires me to be creative
	I always lose track of time in class	I always get to invent my own ways to solve class problems		My teacher always recognizes me for creative work
Creativity is Encouraged	The work I do in class is usually fun	I usually rethink or reframe class problems	I often need to think deeply about class problems	My teacher encourages me to be creative
	I often lose track of time in class	I often get to invent my own ways to solve class problems		My teacher often recognizes me for creative work
Creativity is OK	The work I do in class is sometimes fun	I sometimes rethink or reframe class problems	I regularly need to think about class problems	My teacher allows me to be creative
	I sometimes lose track of time in class	I sometimes get to invent my own ways to solve class problems		My teacher sometimes recognizes me for creative work
Creativity is Discouraged	The work I do in class is usually not fun	I rarely rethink or reframe class problems	I usually don't need to think about class problems	My teacher discourages me from being creative
	I rarely lose track of time in class	I usually have to use the teacher's way to solve class problems		My teacher criticizes me for creative work
Creativity is Not Allowed	The work I do in class is never fun	I never rethink or reframe class problems	I never need to think about class problems	I am not allowed to be creative in class
	I never lose track of time in class	I am required to use the teacher's way to solve class problems		My teacher penalizes me for creative work

The Theory of Climate4Creativity

The proposed theory of Climate4Creativity states that, in order to improve the climate for creativity in a public elementary school, school leaders must take focused action to improve levels of schoolwide community and safety, along with classroom level improvement in the general learning environment and the environment for teaching and learning creativity. This

focused action should take place according to a systematic, strategic process of ongoing improvement in the school. Student perspectives on the landscape of this theory in a particular school would be manifested by the interrelationships between the four core constructs defined by the Climate4Creativity Student Perspectives instruments. Figure 16 illustrates these theoretical interrelationships.

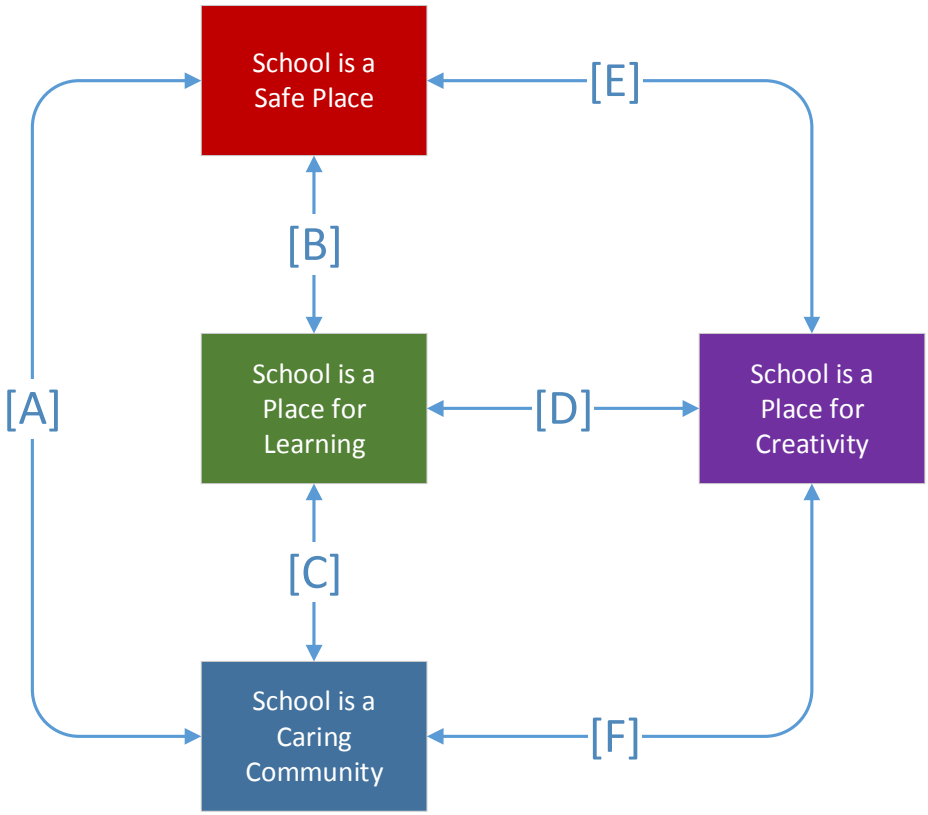


Figure 16. Hypothetical Interrelationships within the Theory of Climate4Creativity.

Applying Theory of Constraints to an Elementary School. In this theory, the four constructs were hypothesized to be highly interrelated, forming a classic system, designed to generate learning. Taking concrete action at any construct would have both direct and indirect effects on the other constructs, and on the performance of the system as a whole. Knowledge of the relative strength of each interrelationship would be essential for development of an

empirically supported *theory of action* to build a climate for creativity in a school, since the potential impact of any tactic is constrained by these interrelationships. From a general systems theory perspective, Goldratt's theory of constraints (Goldratt, 2010), shows that the output from any system is controlled by one or more constraints within the system. By identifying and measuring these constraints, and focusing improvement attention on the most significant constraints at a particular point in time, managers can generate improvement in the output from the system overall. Repeated use of this measure-focus-act cycle forms a systematic, strategic process of ongoing improvement in the system (Goldratt & Cox, 1984, 2004).

What Constrains Teaching and Learning Creativity in Schools? First, safety and community are closely interrelated: For a caring community to exist, feelings of safety are required, and increased feelings of safety are created when a caring community exists within and around a school. The impact of increasing levels of safety on community in an elementary school is roughly equivalent to the impact of increasing levels of community on safety at the school. The combination of high levels of safety and high levels of community enable creation of a positive general learning environment in a school. Improving the learning environment leads to improved feelings of safety and community at school. The effect of increasing levels of safety and community on learning environment is expected to be higher than the effect of improving the learning environment on safety in a school. Without a positive general learning environment, it is impossible to learn creativity, but, learning creativity will in turn lead to an improved general learning environment. The impact of increasing levels of the general learning environment on the environment for creativity is roughly equivalent to the impact of increasing levels of environment for creativity on the general learning environment in a school. Learning creativity will lead to increased feelings of safety and community. These improvements in safety

and community will reinforce the importance of creativity in a school, thereby improving the environment for teaching and learning creativity.

The Theory of Action. This set of hypothetical relations informs a robust theory of action for the climate for creativity, as follows: If our goal is to improve the teaching and learning of creativity in an elementary school, the school must improve the general learning environment. In order to improve the general learning environment, the school must improve levels of safety and community. To improve levels of safety, the school must improve levels of community in the school. In order to improve the school, across these four levers, over the long term, school leaders must measure the current state of the school against the four constructs, focus on the most powerful lever of change at that time, and take effective action to elevate the most powerful lever (and only the most powerful lever). All other constructs will be *automatically* improved, through the focused action taken on the most powerful lever, via the interrelationships between the constructs.

This is a strategic theory of action, taking account of both the most powerful lever at any particular time, and the ability of the human beings within the school organization to execute tactics against this lever. Development of a climate for creativity in a school is a long term, transformational strategy to improve all aspects of a school, potentially requiring multiple years of managed change from successive school leaders. Figure 17 illustrates this theory of action, in hypothesized action at a school.

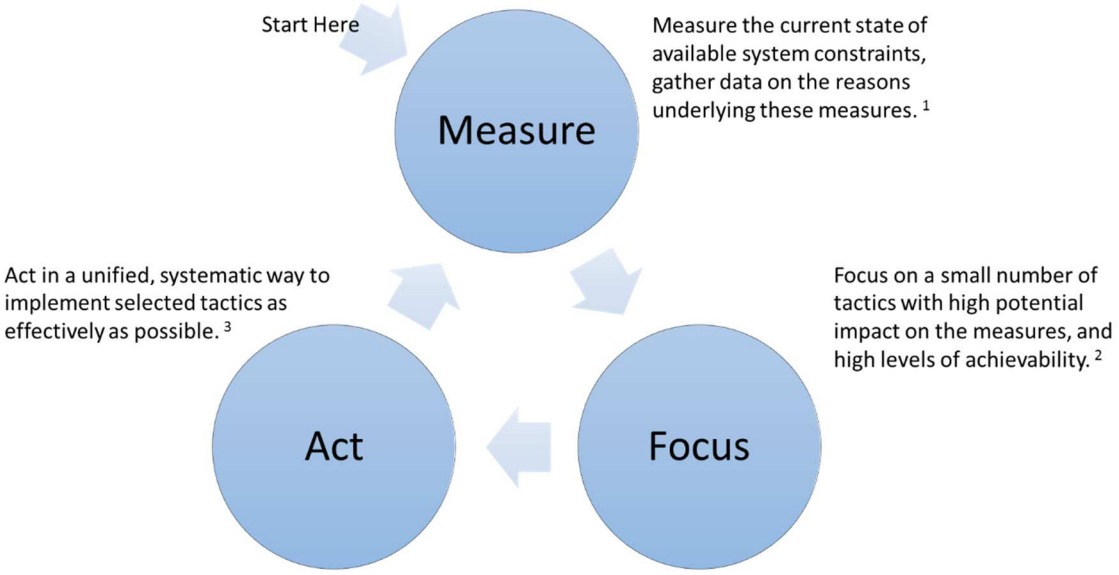


Figure 17. The Theory of Climate4Creativity Measure-Focus-Act Cycle.

[1]Measures will change over time, must be repeated periodically, and will likely call for different tactics at each measurement cycle. [2]It will always be more effective to focus on fewer tactics (ideally a single tactic) at any point in time. [3]One action cycle will never be enough to optimize a system, repeated systematic action is required to improve a system over the long term.

CHAPTER FIVE:

UNDERSTANDING TARGET RESPONDENTS

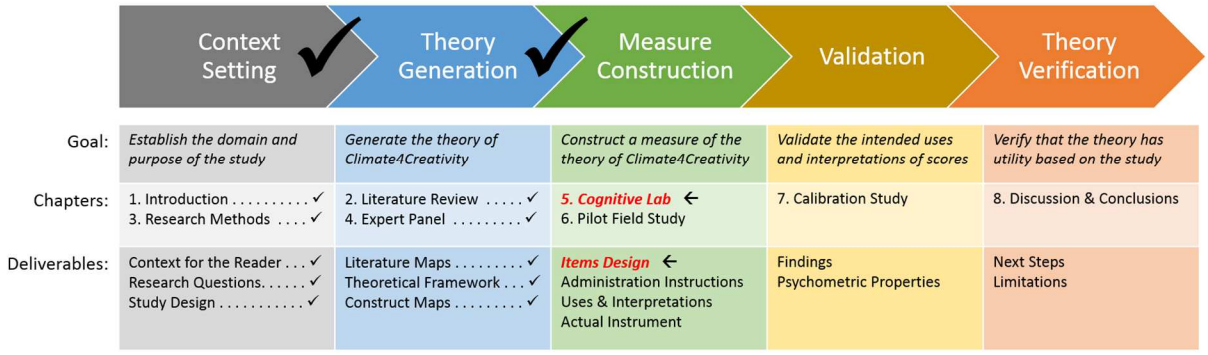


Figure 18. Dissertation Roadmap – Cognitive Lab.

By enhancing the practice of school improvement with a specific focus on *teaching and learning creativity*, an opportunity emerges to better prepare young people to participate in the social life of the planet, thereby empowering the sustainability of human society. To achieve this, an action-oriented *theory of climate for creativity* was proposed in chapter 4. A measure of the theory will be constructed, and the intended uses and interpretations of its scores validated.

Study II: Cognitive Lab

This study was the second phase of a mixed methods research program designed to explore school climate and the environment for teaching and learning creativity in an elementary school setting. This study was designed to provide target respondent input to the following research questions:

RQ1: what are the core elements of healthy school climate & the environment for teaching and learning creativity in a public elementary school setting?

RQ2: how would we measure the degree to which elementary school students believe these core elements exist in their elementary school?

RQ3: what evidence of validity should be collected and what validation arguments should be applied to ensure credibility of such a measure?

RQ4: how might such a measure be employed to actually improve the environment for teaching and learning creativity in a public elementary school?

Introduction. The process used in this study was to pretest the new Climate4Creativity Student Perspectives measurement instrument, developed as an output from Study I of this program of research, with a representative sample of elementary school students drawn from a typical school in the target school district. The pretest took the form of a cognitive lab, employing primarily *think aloud* techniques (Willis, 2005) to explore the cognitive processes used by respondents as they read and answered the items in the instrument. This pretesting process allowed the researcher to identify challenging and ambiguous wording and phrasing, from the target respondents' perspectives, and to trial components of the design with a small sample, in a way that allowed the researcher to learn about the sample and improve the items design in advance of scaling the measurement instrument to a larger field study. In addition, this approach allowed the researcher to investigate both emergent and known issues related to the constructs the instrument was intended to measure.

Participants. Eighteen grade 3 through 8 students, attending Geraldine W. Johnson Elementary school in Bridgeport, CT were interviewed in a cognitive lab format. Johnson school is a typical, midsize urban neighborhood public school serving about 800 students in grades K-8. Six participants were 3rd grade students, three were 4th graders, five were 7th graders, and four were 8th grade students. Eleven students were girls and seven were boys. Seven students were

black, nine were white, one student was Asian, and one Native American. Ten students claimed Hispanic/Latino status.

Procedures. Written approval to work in the Bridgeport schools was provided by the superintendent and board of education. I then contacted the Principal of Johnson Elementary School and she provided written site approval for the cognitive lab to be held on school premises during the regular school day. Once IRB approval had been obtained, a letter was sent home by the principal explaining the research and including an informed legal consent form, addressed to all 3rd, 4th, 7th, and 8th grade parents, asking for parental consent for children to participate in the research. Families were offered $10 Subway restaurant gift cards as an incentive for their child's participation in the study. Twenty-seven consent forms were received by the social worker at the school. Five of the response forms had been completed but were not signed, and a further four did not give explicit permission for audio recording. The remaining eighteen participants were scheduled for 40 minute cognitive lab interviews during the school day, organized into four waves of 4-6 students, spread over a period of four weeks. A major snow storm closed the school for ten days in the middle of the scheduled interviews, extending the cognitive lab period by three weeks, and an additional site approval letter was written by the principal to allow the continuation of the study into March 2013. Informed consent and approval letters for this study can be found in Appendix G.

The cognitive lab was set up in the social worker's office at the school, and the social worker was present in every interview. The lab was set up on my laptop computer, and audio recorded through the computer's internal microphone. Students were sent to the lab by their teachers, according to the predefined schedule, and sat facing the computer to complete their cognitive lab interviews. First, after introducing myself to the student, I explained the conduct of

the lab, and notified students that they could stop the lab at any time, just by saying "stop." The social worker was introduced to them, if they did not already know her. Students were then asked if they were ready to begin. Once their assent had been given, the lab began. At the completion of each lab interview, students were provided with a thank you card addressed to the consenting parent who had signed their informed legal consent form, and containing my business card, an invitation to email me with any follow up questions, and their $10 Subway restaurant gift card.

Data Collection. Data collection was conducted according to cognitive interviewing practice, as documented in Willis (2005). Students were individually shown a variety of multi-choice items on the computer screen, showing both the questions and answer choices, and were asked to read the item and select their answer. They were then asked to think aloud about why they selected the particular answer choice. I used a variety of cognitive strategies to support subjects in articulating their cognitive processes. For example, subjects might be asked to repeat the question in their own words, or to explain the meaning of important words or concepts in the items. In addition, expansive probes such as, *tell me why you selected this answer choice?* were used. These probes served a dual purpose: to both help subjects think aloud, and; to gain a deeper understanding of subject perspectives on important content within the measure (Willis, 2005). Specific expansive probes related to the respondents' definitions of *safety, community, support, bullying*, *creativity*, *kindness, fun, respect, focus*, and *diversity*, were used to explore constructs underlying the Theory of Climate4Creativity from a student perspective. Every interview was recorded using a digital recording device, and digital audio files were saved along with the exact item set reviewed by each student. Note that, in order to ensure full review of all the items, given the limited time available with each respondent, students in the earlier waves of

interviews were asked to respond to a subset of full items design. Students in the final wave responded to the full items design.

Data Analysis. Following completion of each wave of 4-6 interviews, a careful review of the digital audio recordings from each interview was conducted, and adjustments were made to items and answer choices, in an attempt to clarify and simplify the wording of items, and clear up disconnects between the intended interpretation of each item and the actual interpretation of the item. Amended items were subsequently retested in later waves of interviews, until all the major misunderstandings and clarifications no longer appeared in the final wave of interviews. A total of about nine hours of audio recordings were analyzed. Expansive probe responses were informally categorized by topic and used to inform wider decisions such as how to provide an engaging but unbiased introduction to the instrument, when to provide definitions of words, how many answer choices were appropriate for each age group/developmental level.

Results. Findings from this study were ultimately categorized into four groups of insights – demographics, caring community, safety/bullying, and creativity/learning. The findings for each category are explored below.

Category 1 – demographics. Several Hispanic/Latino students were not familiar with the words "Hispanic/Latino," when probed, they identified themselves as Spanish. Several students commented that they were of mixed race, struggling to select a single racial option from the list. Third graders particularly struggled with racial identity questions, 4th graders and above seemed to understand. The phrase *Native American* particularly caused confusion in native Spanish speakers, who paraphrased this category as, *someone born in America*. The word *gender* was confusing to younger students, two of whom were unable to paraphrase the word. When probed,

it seemed that they understood the question by reading the answer choices (*male* and *female*), which they paraphrased into *boy* and *girl*.

Category 2 - caring community. One student commented that she only had one parent, and two commented that their grandparents took care of them at home, the phrase *report card conference* was easier to understand than *parent teacher conference*. Several students felt that their family members *checked* their homework, rather than *helped* with their homework. They viewed these as different activities. One third grader struggled to interpret the difference between *closest friends* and *other students in my class*, but other students were able to easily differentiate. The word *kind* was interpreted to mean the opposite of *mean*, and *respectful* to mean the opposite of *rude*. Most students in the study claimed to like or love their school. Expansive probes around what it would take to make students hate or dislike their school solicited responses such as people being mean or rude, and a variety of comments about specific programs not being in place (for example, big brothers big sisters, and the school's lighthouse after school program). Only one student had ever heard of a *parents' organization* at their school (although Johnson school has an active parents' organization in place). One expansive probe led to the word *drama* being mentioned several times, particularly by older students, in relation to how people treat each other. Students in subsequent waves could articulate what this meant, in the context of how students interact at school.

Category 3 – safety/bullying. Across the age ranges in the interviews, students could easily determine if something met the definition of *bullying* behavior, the only significant misunderstanding for students was the belief that bullying did not include harm or theft of their property. Interestingly, several students felt that the answer choices, *encourage the fight* or *stand around and watch* were important missing options in response to the question of what students

do when they see bullying at school. Respondents struggled to pick one option from the choices provided. The phrases *completely safe* and *completely unsafe* were somewhat challenging for several students, they felt that school could never be completely safe or unsafe. The word *corridor* caused confusion for several students, they paraphrased it as *hallway*. Older (grade 7 and 8) students reported that they did not have recess, in fact they had no outside time during the school day. None of the students in the study traveled to school by school bus, so they could not complete the school bus questions. Bathrooms were seen as dangerous places both from the perspective of being out of sight of teachers, and as places with hygiene issues, which might lead to getting sick. Probes around the role of the security guards at school showed that students in the study felt the security guards job was to watch students rather than to keep them safe from outside threats. Finally, the word *drama* came up repeatedly in questions about safety and bullying. It seemed to be a pre-cursor to acts of bullying, and difficult for adults to tell the difference (although the students themselves claim to know the difference).

Category 4 – creativity/learning. The word *creative* had a variety of definitions, one respondent claimed it was what teachers call you when you don't know the answer to a question. Most students used art or music or sports when asked to give examples of what creative meant to them, one respondent discussed at length the extensive use of glitter in creative activities. Students felt that teachers always encourage creativity, but that hard work and learning was much more important than being creative. Students did not see creativity as something learned, but as an inherent trait – you either are creative or are not. Most students connected creativity with fun. All teachers give clear directions, and most expected excellent or good work from the students in the study. One student said he didn't know what his teacher expected, because the teacher never spoke to him. Questions about having time to finish their work confused several

students – the practice at this school was that any work not finished in class is done as homework. The question of variety of activities in class confused several students, they were unable to separate different subjects from different activities. Two older students asked which class they should think about when answering questions about learning and creativity.

Discussion. Based on early cognitive lab interviews, a number of changes were made to the measurement instruments. These changes were retested with later waves of interviews. These changes are discussed in the four categories of findings.

Category 1 – demographics. In order to address respondent perspectives on demographics, the Hispanic/Latino identity question was moved to *before* the racial identity question and separated from the question of what language is used at home. This change served to reduce the confusion around Spanish and Hispanic/Latino wording. The gender question was changed to, *are you a girl or a boy?* to simplify the choice for younger students. The racial identity question was moved to the end of the demographic questions and amended to allow selection of multiple racial groups. In addition, the racial identity answer choices were changed to: *White (European)*, *Black/African American*, *Native American (American Indian)*, *Asian American*, *Pacific Islander*, and *Other Race*. This change clarified the meaning of *Native American* for younger Hispanic/Latino students. Both elementary and middle school instruments were changed to use exactly the same wording for all demographic questions.

Category 2 – caring community. In order to clarify wording, the phrase *family members* will be used throughout the measure, rather than *parents*. Some students suggested that family members *check* their homework rather than *help* with their homework, this led to a clarification of family support questions, with *helping* and *encouraging* and *talking to teachers* being

separately asked. The question about parents' organizations was removed from the elementary measure, but maintained in the middle school measure.

Category 3 – bullying/safety. The word "completely" was removed from the safety scale, which became, *Very Safe*, *Safe*, *Mostly Safe*, *A Little Unsafe*, *Very Unsafe*. The word *corridor* was replaced everywhere by the word *hallway*. Questions about bullying were changed to ask questions at a more granular level first, intrinsically defining bullying for students. For example, *how well does your school keep you safe from being hurt? how well does your school keep your stuff safe? how well does your school teach you about bullying?* became early questions asked in advance the question, *how often do you see people bullying others at your school?* One important finding contradicts informal discussions with teachers and school leaders: students as young as grade 3 *do* have a clear understanding of what bullying means, and, in fact have a *lower* standard than the Connecticut legal definition of bullying, particularly with respect to damage and theft of personal property (An Act Concerning the Strengthening of School Bullying Laws of 2011).

A specific question about drama at school was added to the middle school survey, along with questions about people being mean and people bullying others. The questions about safety at recess was removed from the middle school measure. Finally, instead of questions related specifically to the role of security guards, security items were changed to focus on how well the school stops people coming in rather than what the security guards do.

Category 4 – creativity/learning. Creativity and general learning environment questions were intermingled within the items design to provide a clearer flow for the respondent, rather than being viewed as separate categories of questions. The middle school measure was changed to add two questions near the beginning of the survey – to identify the favorite and least favorite

subject. All learning environment and creativity items were then asked twice, once for the favorite and once for the least favorite classroom. This change served to eliminate confusion about which classroom respondents should consider in answering learning and creativity questions. Learning expectations questions were improved by adding a self-expectations question (*what grade do you expect to earn in class?*) for middle school students, and the *time to finish my work* question was split into two questions, in an attempt to clarify the issue of classroom time management versus losing track of time.

Conclusions and Limitations. This study provided excellent input into the design of specific questions and answer choices to be used in the elementary and middle school measures of the climate for creativity, as well as providing input into the overall structure of the measures. The study was limited in scope, in that only 18 students were interviewed, making findings difficult to generalize. In addition, the students interviewed were all *good* students at Johnson school. It is possible that a wider sample or a random sample of students would have had more trouble with the questions, particularly at the lower ages, leading to a richer series of refinements.

CHAPTER SIX:

PILOTING THE INSTRUMENT

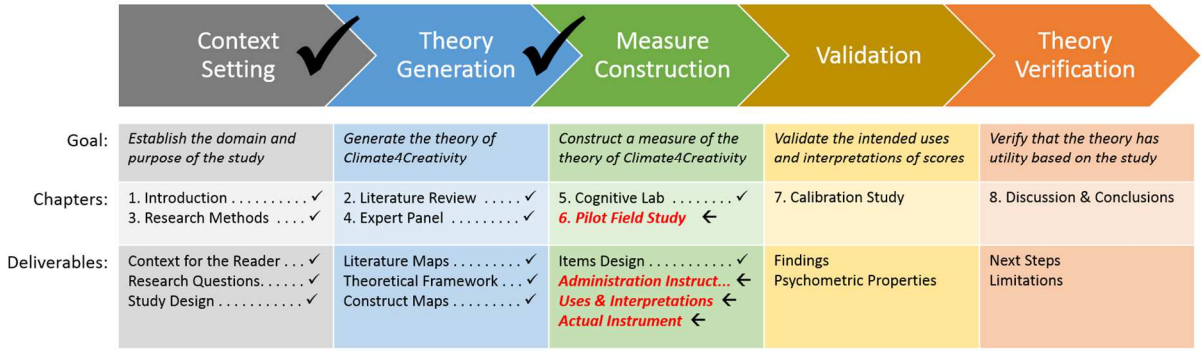


Figure 19. Dissertation Roadmap – Pilot Field Study.

By enhancing the practice of school improvement with a specific focus on *teaching and learning creativity*, an opportunity emerges to better prepare young people to participate in the social life of the planet, thereby empowering the sustainability of human society. To achieve this, an action-oriented *theory of climate for creativity* was proposed in chapter 4, a measure of the theory constructed, and the intended uses and interpretations of its scores validated.

Study III: Pilot Field Test

This study was the third phase of a mixed methods research program designed to explore school climate and the environment for teaching and learning creativity in an elementary school setting. This study was designed to provide initial, relatively large quantitative input to the following research questions:

RQ1: what are the core elements of healthy school climate & the environment for teaching and learning creativity in a public elementary school setting?

RQ2: how would we measure the degree to which elementary school students believe these core elements exist in their elementary school?

RQ3: what evidence of validity should be collected and what validation arguments should be applied to ensure credibility of such a measure?

RQ4: how might such a measure be employed to actually improve the environment for teaching and learning creativity in a public elementary school?

Introduction. The process used in this study was to pretest the new Climate4Creativity Student Perspectives measurement instruments, developed as an output from Study I of this program of research, with a representative sample of elementary school students. This pretest took the form of a complete fielding of the instrument in grades 3-8 at one elementary school in Bridgeport, CT.

Participants. A total of 285 students attending High Horizons Magnet School in Bridgeport, CT completed the version 1.1 instruments during February and March 2013. High Horizons was a district magnet school located in the heart of the city of Bridgeport, and catering to about 500 students in grades PK-8. All students were drawn from the local Bridgeport neighborhoods and gained entry to the school via lottery. Most students (94%) in grades 3-8 completed the pilot instrument: 148 students completed the elementary survey (grades 3-5, v1.1e) and 137 completed the middle school age survey (grades 6-8, v1.1m). In addition, a cohort of six 6th grade students with special needs completed the elementary (grades 3-5) survey and were included along with grade 6 data. Seventy six (76) respondents were white, 130 were black/African American, and seventeen (17) were from other races. Ninety-eight (98) students

claimed Hispanic/Latino status. Figure 20 shows a breakdown of pilot study respondents by grade level and gender.

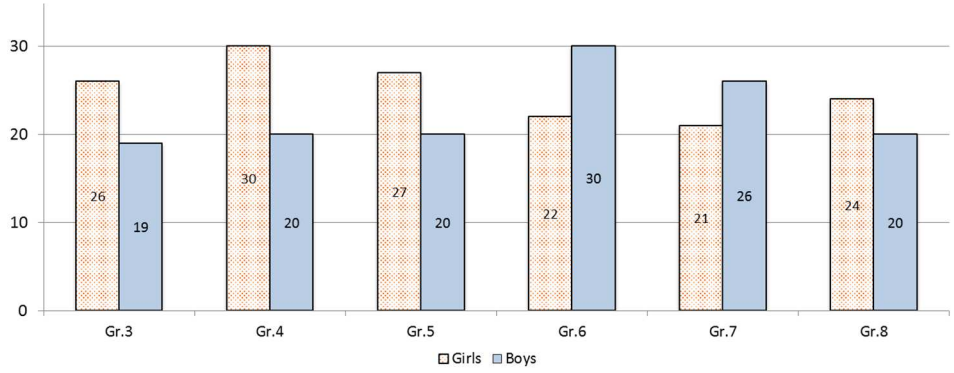


Figure 20. Pilot Respondents by Grade Level and Gender.

Procedures and Materials. Written approval to work in the Bridgeport schools was provided by the superintendent and board of education. I then contacted the principal of High Horizons Magnet School (HHMS), and she provided written site approval for the study to be held on school premises during the regular school day. Once IRB approval had been obtained, a letter was sent home by the principal explaining the research and notifying parents that the study was taking place during the normal school day. I worked directly with the school's computer literacy teacher to organize administration of the instrument in the school's computer lab and visited the school on two occasions to test the instrument and to train the computer literacy teacher on the conduct of the study. I was not present on school grounds during the collection of data, but checked in with the computer literacy teacher by phone and email periodically during the conduct of the test.

Students cycled through the computer lab during their normal scheduled computer lab time, and completed the instrument in the lab along with their classmates. Clear instructions were read to students by the computer literacy teacher. Students were informed that they did not

have to complete the survey and that their answers would be kept confidential. In addition, individual informed assent was obtained from students on the first page of the electronic survey. On completion of the survey, students remained in the computer lab, and the computer literacy teacher conducted a lesson on the use of computer surveys, before preparing the computers for the next scheduled class of students. If students were absent on the day of administration for their class, they were not given an opportunity to complete the survey at another time. The complete instrument (v1.1) used for the pilot study is provided in Appendix D, and legal consent documentation can be seen in Appendix G.

Data Collection. Students completed an online instrument, version 1.1. The instrument was delivered to students using the SurveyMonkey online survey tool. Students in grades 3-5 completed the elementary instrument, and students in grades 6-8 completed the middle school instrument, with the exception of six 6th grade students with special needs, who were directed by the computer literacy teacher to complete the elementary instrument. The elementary survey used primarily four point scales for each item in the items design, and the middle school survey used primarily five point scales.

Data Analysis. The purpose of this pilot study was to further refine the items design for the instrument and to gather evidence of validity of the scores yielded by the instrument for its intended uses and interpretations. Data from the two pilot instruments were extracted from SurveyMonkey into two separate Microsoft Excel (Microsoft, 2013a) spreadsheets, and reformatted into a text format readable by the Winsteps (Linacre, 2014) software tool. Four separate Winsteps control files were created for each data file (eight control files in total), using the IDELETE= command to enable analysis for each construct separately, using the same underlying data files. Within each instrument, five analytical exercises were conducted for each

construct, designed to identify suspect items within the items for that construct. Suspect items were then further investigated using a variety of analytical tools, and qualitative refinement techniques. Note that where I refer to an item or an answer choice in this section, I will show the item code in plain text and the item text in italics. A full definition of the pilot items and answer choices can be found in Appendix D.

Pilot Reliability Analysis. Reliability is a measure of the consistency of responses to an instrument across different administrations. Person raw score reliability was computed using Cronbach's Alpha [KR-20], along with a Winsteps (Linacre, 2014) modeled, error-adjusted person separation reliability statistic. A modeled reliability of 0.8 or higher is recommended to provide an instrument that will reliability distinguish between high and low performance on the measure being examined (Linacre, 2012). The most effective way to increase the reliability statistic of an instrument is to increase the number of well-fitting items in the instrument.

Item Polarity Analysis (point-measure correlations). A fundamental concept in Rasch measurement is, when a person has a higher latent belief or feeling about a construct, they should pick a higher level on the items in the measurement instrument; and, that a higher level on the items therefore implies a higher measure in the person (Bond & Fox, 2007). Item polarity analysis examines the relationships between the person ability (or in this case, the degree or level of feeling about an item) and the item difficulty, to establish if the category scale for a particular item is oriented properly. Essentially, this analysis answers the question: *do the responses to this item align with the feelings of the persons who answered it?* Winsteps provides a statistic called the empirical point-measure correlation (PT-CORR), which examines the extent to which responses to items correlate to overall measures for the respondents involved. Clearly, negative correlations (between -1.0 and 0.0) indicate that an item's response scale is oriented in the

reverse direction to the overall measure. Small positive correlations (between 0.0 and 0.3) indicate that further examination of the item wording or answer choices may be required, moderate positive correlations (between 0.3 and 0.7) indicate good item polarity, and high positive correlations (between 0.7 and 1.0) indicate excellent item polarity, implying very strong relationships between the levels for a particular item and the latent feeling of the persons responding to the item. Any items identified as having PT-CORR values of 0.3 or less will be identified as suspect items for further analysis, items with PT-CORR values above 0.3 are assumed to have appropriate item polarity for the purposes of measurement using the instrument being tested.

Analysis of Empirical Item-Category Measures (item-person maps). This analysis allows the researcher to examine the items design overall, and investigate the validity of the construct maps, and the distribution of the measures of respondents in the pilot sample, through the use of item-person maps. This is the closest practical analytical tool to an actual pathways analysis, described in the introduction to this chapter. The Winsteps (Linacre, 2014) software tool provides item-person maps, designed to allow comparison of the distribution of difficulty of items in the instrument, with the distribution of respondent levels on the measure. This allows the researcher to qualitatively examine the vertical hierarchy of items, and see if the items design is actually answering the questions it is intended to answer. In addition, if the instrument is valid for its intended purpose, it should provide a fairly even distribution of items on the item-person map, and a fairly even distribution of persons horizontally on the item-person map.

Category Distractor Analysis. It is important in a well-designed instrument, that the answer choices for each item are sequenced so as to ensure that a higher level on the nominal scale equates to a higher level of the latent variable. Suspect item categories can be identified on

item-person maps, and through the detailed item polarity data tables provided by Winsteps (Linacre, 2014). Items where categories order is not as expected are identified for further analysis with an "*" on the data tables, and are shown in the wrong order on the item-person maps. These items can then be investigated in more detail by examination of empirical and expected item characteristic curves. These curves visually compare the expected item characteristics from the Rasch measurement model with the empirical data from the pilot study, giving qualitative clues to help refine the items and answer choices in the instrument.

Item Fit Statistics. The Rasch model asserts that a person of an average level on a measure will likely answer easier items at a high level, hard items at a low level, and moderate items at a moderate level. In other words, an average person should answer easy items correctly, hard items incorrectly, and get moderate questions right about half the time. Because we understand the level of difficulty of our items and the level of ability of our respondents, we can use this knowledge to analyze educational tests for cheating, and, in the context of this pilot test, we can identify items that likely won't fit Rasch model assumptions later, and are therefore suspects for additional analysis. Analysis of item fit statistics in Winsteps, using mean-square fit parameters allows us to examine the fit characteristics of the pilot respondents and items. A mean-square fit parameter of close 1.0 implies that the item is sound for measurement purposes. High mean-square values (above 2.0) distort the measurement system and should be eliminated from an instrument, low mean-square values (below 0.5) do not degrade measurement, but may lead to artificially high reliability measures, and so should be avoided where possible. Rasch measurement transactions (Wright & Linacre, 1994) suggests a parameter range from 0.5 to 1.7 be considered a good fit for a clinical instrument, a range of 0.7 to 1.3 for typical measurement instruments, and 0.8 to 1.2 for high stakes testing instruments. Winsteps provides data about the

fit characteristics of inliers (infit) and outliers (outfit) separately, and allows the researcher to target reasonable mean-square fit values on both items. For the purposes of this analysis, the clinical decision making range of parameters was used, targeting mean-square parameters in the 0.5 to 1.7 range.

Results. The Winsteps software tool was used to analyze the data from the pilot elementary and middle school instruments. Results are presented in Tables 9 and 10.

Table 9.
Pilot Analysis for the C4C/SP Elementary Instrument, v1.1.

Pilot Analysis		School is a Safe Place	School is a Caring Community	School is a Place for Learning	School is a Place for Creativity
Pilot Reliability	Raw Score Reliability[1]	0.92	0.84	0.81	0.82
	Modeled Reliability[2]	0.85	0.77	0.62	0.63
Item Polarity	Items with Negative PT-CORR	12C: -0.07	None	None	None
	Items with PT-CORR ≥ 0 and < 0.3	12B: 0.12 12A: 0.23	6C: 0.07 6B: 0.16	19: 0.24 21D: 0.27	None
Empirical Item-Category Measures	Items with Categories out of Sequence[3]	11C: 2134 13C: 2134 13D: 2134 12A: 2134 12B: 1324 12C: 3412	6A: 2134 8E: 2134 6C: 1324	20C: 2134	None
	Items with Missing Categories	11A: 1 10C: 2	None	19: 1,2	None
Category Distractors	Items with Potential Distractors	12C 12B 12A 13C 13D 11C	6C 8E 9 6A	20C	None
Item Fit Statistics	Items with High Infit[4] or Outfit[5] MSQ[6]	12C In[4]: 2.31* Out[5]: 5.69* 12B In[4]: 2.07* Out[5]: 2.10* 12D In[4]: 1.70 Out[5]: 1.74 12A In[4]: 1.72 Out[5]: 1.64	6C In[4]: 1.34 Out[5]: 2.34* 7C In[4]: 1.71 Out[5]: 1.65	None	None
	Items with Low Infit[4] or Outfit[5] MSQ[6]	None	None	None	None

Notes. n=145. [1]Cronbach's Alpha. [2]Winsteps Error Adjusted Reliability. [3]Should be "1234". [4]Fit characteristics of inliers. [5]Fit characteristics of outliers. [6]Mean-square fit statistic.

Table 10.
Pilot Analysis for the C4C/SP Middle School Instrument, v1.1.

Pilot Analysis		School is a Safe Place	School is a Caring Community	School is a Place for Learning	School is a Place for Creativity
Pilot Reliability	Raw Score Reliability[1]	0.93	0.80	0.83	0.85
	Modeled Reliability[2]	0.92	0.82	0.79	0.83
Item Polarity	Items with Negative PT-CORR	None	None	None	None
	Items with PT-CORR ≥ 0 and < 0.3	14C: 0.14 16A: 0.17 14A: 0.23 14B: 0.23	8B: 0.20	29: 0.02 22B: 0.05 22A: 0.07 21: 0.25	None
	Items with PT-CORR ≥ 0.70	13G: 0.70	None	None	30C: 0.70 30E: 0.71
Empirical Item-Category Measures	Items with Categories out of Sequence[3]	19C: 21345 15C: 21345 15A: 21345 12F: 13245 15E: 23415 14A: 23451 14B: 34*25	11: 21345	29: 21345 26: 23145 24E: 21345 23H: 31245	23F: 21345 23E: 21345 24B: 21345 23C: 21345
	Items with Missing Categories	14B: 1 14C: 1,3	9D: 1 8B: 1,2	28: 2 21: 1 23A: 1 22A: 1,2	25: 2
Category Distractors	Items with Potential Distractors	14A 14B 14D 16B 18D 19C 15E 18A 12E 12F 15A 15C 13A	8F 10D 11 9E	29 22B 27 26 24E 23H 24C	24D 24B 23C 23E 23F
Item Fit Statistics	Items with High Infit[4] or Outfit[5] MSQ[6]	14C In[4] 1.60 Out[5] 1.81 14D 2.04 2.31	8F In[4] 1.76 Out[5] 1.68	None	24D In[4] 1.76 Out[5] 1.88
	Items with Low Infit[4] or Outfit[5] MSQ[6]	None	None	None	None

Notes. n=137. [1]Cronbach's Alpha. [2]Winsteps Error Adjusted Reliability. [3]Should be "12345". [4]Fit characteristics of inliers. [5]Fit characteristics of outliers. [6]Mean-square fit statistic.

Discussion. In general, reliability scores were somewhat higher on the middle school instrument than on the elementary instrument. This is not surprising, since the middle school instrument has more items, and it is likely that older respondents have a more even understanding of the items and answer choices in the instrument. Overall, the standard Cronbach's Alpha reliability scores were acceptable ($\alpha \geq 0.8$) for all four constructs in both instruments, and were particularly high for the School is a Safe Place construct on both the elementary ($\alpha = 0.92$) and middle school ($\alpha = 0.93$) instrument. The modeled, error-adjusted person separation reliability scores provided by Winsteps, are more sensitive to the separation values for elements in a Rasch measure, and showed unacceptably low values for the School is a Place to Learn ($r = 0.62$) and School is a Place for Creativity ($r = 0.63$) constructs in the elementary instrument. The implication of this finding was that there were too few good fit items in these constructs within the elementary instrument. Additional items will be required to increase the modeled reliability for these constructs. The modeled reliability scores for the elementary version of the School is a Caring Community construct ($r = 0.77$) and the middle school version of the School is a Place for Learning construct ($r = 0.79$), were moderate values, close to 0.8, and will likely be improved by refinement of the suspect items found in this analysis, without addition of multiple new items.

Item polarity analysis identified a number of suspect items. The most serious issue was a single negative item polarity on the elementary instrument (polarity = -0.07), implying that the scale processes in the opposite direction than expected. This is a near zero negative polarity, and is therefore likely not a simple misdirected scale. The item, 12C: *People have to buzz in*, is part of a short series of four items designed to measure the degree to which students feel that their school protects them from people coming into the school. This series is the primary measure of

security from outside threats – a core subconstruct within the safety construct. One of the other items in this series, 12B: *visitors have to wear stickers,* also had a low positive polarity (polarity = 0.16). All four items (12A-D) have poor infit and outfit characteristics, with 12C and 12B having the worst fit with the model. It is notable that three of the comparable middle school items also had unacceptably low, although positive, item polarities (items 14C: 0.14; 14A: 0.23; 14B: 0.23), and two of the four comparable items on the middle school scale also have poor fit characteristics (14C, 14D). The implication of this evidence is that this particular item series may not be part of a true measure of feelings of safety at a school, from a student perspective. Students may not clearly associate these four items regarding stopping people coming in with their feelings of safety at a school. While it is important to have security at the door, this finding would imply that having door security or not bears little relationship to student feelings of safety in the school, this corresponds to input from the cognitive lab participants in Study II who generally felt that security guards were there to police student activities, not to protect students from outside threats.

Within the community construct, two items in the elementary instrument were identified as having low item-polarity characteristics (6C: 0.07; 6B: 0.16). One corresponding item on the middle school instrument (8B: 0.20) also showed a low item-polarity characteristic. These items are related to a five item series on how family members are involved with the school. Item 6B/8B: *they read my report cards*, appears to be so simple to answer that it is not a relevant part of an overall measure of community. Essentially, virtually every student has a family member who reads her report card, making this item irrelevant as part of a measure, this is further evidenced by missing response categories 1 and 2 on the middle school instrument: not a single middle school respondent selected *never* or *not often* on this item. These items do, however, fit

the model and potentially provide some useful data for the occasional student whose family do not, in fact, read her report card. Item 6C: *they tell me to work hard*, also showed a low item polarity in the elementary instrument (0.07), and was not a good fitting item for outliers in the main response pool. The word, *tell*, used in the elementary instrument, but replaced by the word *encourage* in the middle school instrument looks likely to have changed the underlying meaning of the answer choices for this item, for young respondents. I suspect that some young students interpreted the word *tell* in a negative way, thereby reversing the polarity of the scale for those respondents, and leading to poor outfit characteristics. These poor fit characteristics were not observed in the middle school item. The elementary school item will therefore be changed to use the word *encourage* instead of the word *tell*. A related item on the elementary instrument, 6A: *they help me with my homework*, showed as a category distractor, with category 1: *Never* and category 2: *Sometimes* misordered. Since this item showed good model fit, and there was no comparable evidence of distractor issues from item 8A on the middle school instrument, it is likely that this is another item similar to item 6B, where few students will say their families never help them with their homework.

In the learning construct items, item 19: *what does your teacher expect from you?* was identified as having poor item-polarity characteristics, primarily because option 1: *Nothing* and option 2: *Bad Work* were not selected by any students in the pilot. The comparable item in the middle school instrument showed exactly the same characteristics for favorite subject selections, although the options were 1: *expects me to fail* and 2: *poor work*. Middle school item 28: *What does your [least favorite subject] teacher expect from you?* also had a missing category, option 2: *poor work* was missing, but this item did not show poor item-polarity. Middle school item 21: *what grade do you usually earning in [favorite subject] class?* showed similar results, with

option 1: *F* missing from the data. No student feels that they usually earn an F in their favorite class. The implication of these results is that this series of items are appropriate, and might, in fact, demonstrate that teacher expectations really are related to student motivation.

Elementary survey item 21D: *the work I do is difficult*, showed a slightly low item-polarity (0.27), and the equivalent middle school items, 22B/29: *How difficult is the work you do in [favorite subject/least favorite subject] class?* showed a similar low item-polarity (0.05 and 0.02 respectively). This issue is likely related to the nature of the question as part of a measure of the learning environment. Does more difficult work, or the feeling of more difficulty in the work, really mean a better learning environment, from an elementary student perspective? The assumption in the scale is that a better learning environment includes more challenging work. The data from this pilot does not support this assumption, particularly in the middle school instrument, where this is little relationship between the direction of these items and the overall measure. The elementary item has a stronger potential relationship, a polarity of 0.27 is close to the target of 0.3, so could continue to be used as part of the learning construct measure. As a precautionary measure, these items should be removed from the measure, but maintained as interesting detail items in the instrument.

The final suspect item based on item polarity, was middle school item 16A: *people talk about bullying* (polarity = 0.17). It is possible that this item was interpreted by some students to imply teachers talk about bullying rather than other students talking about bullying, leading to a misdirection or distraction to the measure. By rewording this item in terms of *students talking about being mean* rather than *people talk about bullying* this issue should be eliminated from the next version of the instrument.

Several items were flagged as suspect items through both category distractor analysis and empirical item-characteristic measures. Elementary safety items, 11C: *going to the bathroom*, 13C: *they help me not be afraid at school*, and 13D: *they make school a nice place*, all showed misordered category responses, with the categories 1 and 2 out of sequence in the pilot responses. The similar middle school items, 15A: *they keep me safe from getting hurt*, 15C: *they help me not be afraid at school*, and 15E: *they teach us about bullying* show similar misordered responses. In addition, six other elementary items showed as potential distractors. In each case, further investigation showed that these items were likely identified due to missing response categories, and so were not changed based on this analysis. One additional elementary item, 7C: *students in other grades help me*, showed marginal fit characteristics (infit = 1.71; outfit = 1.65). Since the target range for this analysis was 0.5 to 1.7, it was determined that this was close enough for this stage, and further examination of this item would be completed with a broader population. Similarly, one middle school item, 8F: *they are involved in a parents' organization* showed marginal fit characteristics (infit = 1.76; outfit = 1.68). In this case, it is likely that many students were unaware of the answer to this question. It was therefore removed as a component of the community measure, but maintained as part of the measure overall as a useful detail item.

Conclusions. First, the security from outside threats construct will be removed from the overall measure of student feelings of safety, but will be maintained as a series of detailed items. It is proposed that adults in the school will likely associate security protections with feelings of safety, and potentially, older students (upper middle school and high school students) may carry this association. In the future, it is likely that a measure of school staff and family perspectives will be constructed, and these items will provide meaningful comparators for these later

instruments. For elementary students, however, there is evidence that this subconstruct does not contribute meaningfully to the measure, and therefore, it will be removed from the measure and from the student perspectives part of the theory of Climate4Creativity.

Second, a number of wording changes were implemented, as discussed above. In addition, all measurement scales will be adjusted to use five-point scales rather than four-point scales, and the wording will be generally synchronized between the elementary and middle school instruments. This synchronization served to make the instruments more consistent and provided for the possibility of combining data from the two instruments at a later stage, following a subsequent equating study. Several new items were added to increase reliability of the creativity and learning constructs. A small number of items were maintained in the instrument, but disconnected from the construct measures.

The Completed Items Design

Based on the findings from both Study II (cognitive lab) and Study III (pilot test), a number of structural adjustments were made to both the elementary and middle school items designs. Primary amongst these was the decision to move both instruments to identical item wording and identical 5-point scales, where practically possible, and the decision to remove the security from outside threats subconstruct from the overall measure of school is a safe place. Several individual items were kept in the items design, but disconnected from the construct and overall measures. This decision was intended to maintain the actionability of the data yielded by the instrument, by providing details not included in the measures, while simultaneously maintaining the psychometric qualities of the measures. The full completed items design, known as version 1.2 of the instruments, can be found in Appendix E of this document.

Intended Interpretations and Uses of the Measure

It is important to clearly document the intended uses and interpretations of a measure as part of the design and pretesting process. In this case, the Climate4Creativity Student Perspectives instruments are intended for one major purpose and two secondary purposes. The major purpose of the measure is as a benchmarking and focusing tool for school improvement, to enable a school principal to orient herself to the barriers to improvement in a school, and to focus attention on the most important areas to take action to improve the school. The secondary purposes include providing input to tactical decision making at the classroom and schoolwide level, and providing external data for accountability reporting to state entities and the public. Figure 21 illustrates the intended uses and interpretations of the Climate4Creativity Student Perspectives (C4C/SPE and C4C/SPM) instruments.

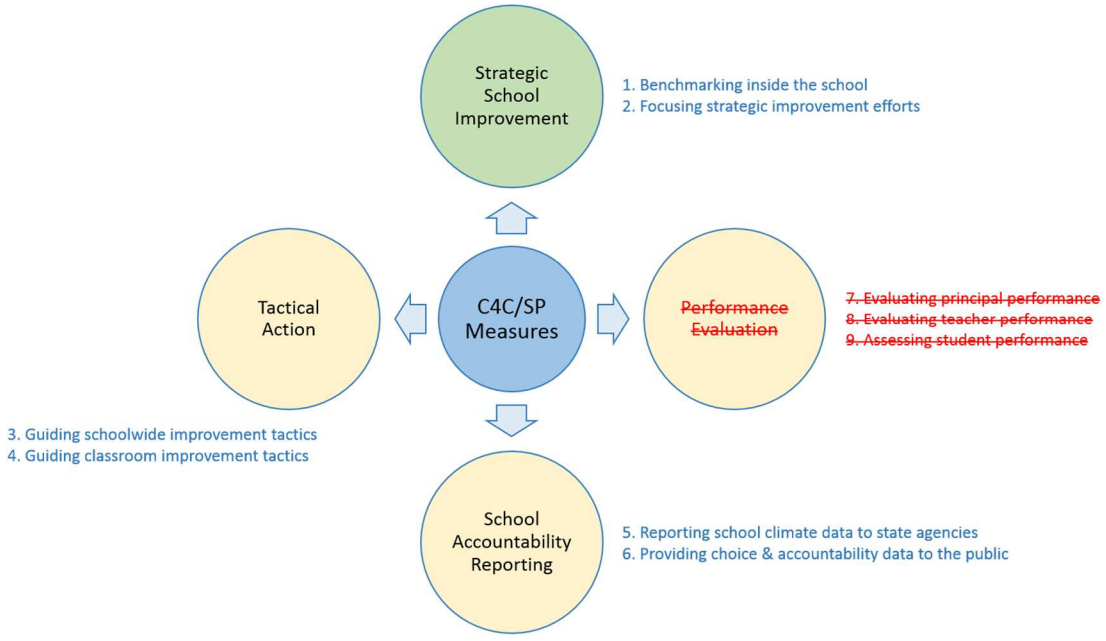


Figure 21. Intended Uses and Interpretations of the C4C/SP Measures.

Notes. Green circle denotes primary purpose, yellow circles denote secondary purpose. Items 7, 8 and 9 are included to clearly articulate inappropriate uses of the instruments.

Strategic School Improvement. The measurement scores yielded by the Climate4Creativity Student Perspectives (elementary) instrument (C4C/SPE) and Climate4Creativity Student Perspectives (middle school) instrument (C4C/SPM) are intended to provide actionable data to help school leaders (superintendents and principals) improve their schools in strategic ways.

1. BENCHMARKING INSIDE THE SCHOOL: Taking action to drive sustained improvement in the overall climate for creativity in a school is a *primary intended use* of the measure. The overall measure and construct measures may, therefore, be used as an annual benchmark for a school, to show improvement or deterioration in the creative climate.
2. FOCUSING STRATEGIC IMPROVEMENT EFFORTS: The construct measures provide a snapshot of the levels of the levers of school change, according to the Theory of the Climate4Creativity, described in chapter 4 of this document. These snapshots are intended to be used to inform school leaders and help them focus on the primary constraints to improvement of their school, and thereby to focus attention on actions needed to improve performance at the most constrained construct, in the context of a defined schoolwide improvement strategy. This is a *primary intended use* of this measure.

Tactical Action. The measurement scores yielded by the Climate4Creativity Student Perspectives (elementary) instrument (C4C/SPE) and Climate4Creativity Student Perspectives (middle school) instrument (C4C/SPM) may be used to provide actionable data to help school leaders (superintendents and principals) and teachers improve their schools in tactical ways.

3. GUIDING SCHOOLWIDE IMPROVEMENT TACTICS: Responses to individual items within the instrument (provided the sample size is sufficient for the individual items to provide statistical significance), may be used to guide decision making about specific tactics and interventions to be employed at a particular construct, according to the Theory of the Climate4Creativity, described in chapter 4 of this document.
4. GUIDING CLASSROOM LEVEL IMPROVEMENT TACTICS: Responses to individual items within the instrument (provided the sample size is sufficient for the individual items to provide statistical significance), may be used to provide general guidance to teachers about tactics and interventions to improve the learning environment in their classrooms.

School Accountability Reporting. The measurement scores yielded by the Climate4Creativity Student Perspectives (elementary) instrument (C4C/SPE) and Climate4Creativity Student Perspectives (middle school) instrument (C4C/SPM) may be used to provide accountability data to state agencies (as required by school climate and antibullying legislation) and may be published at the overall measure and construct measure level to provide accountability data for public consumption.

5. REPORTING SCHOOL CLIMATE DATA TO STATE AGENCIES: The safety and community measures, along with specific items within each measure (provided the sample size is sufficient for the individual items to provide statistical significance), may be used by schools to report antibullying and school climate data to state departments of education and other outside agencies.
6. PROVIDING CHOICE & ACCOUNTABILITY DATA TO THE PUBLIC: The construct measures provide a snapshot of the learning environment and culture at a

school, and may be used to help families make decisions about which schools they would like their children to attend. This may be presented to the community as a league table or comparative report designed to support parental choice and school accountability. Appropriate data quality and sample size information must be provided with any public reporting of this data.

Inappropriate Use of the Measures. It is important to note that the findings from the Climate4Creativity Student Perspectives instruments should not be used for the purposes of principal evaluation, teacher evaluation, or individual student assessment. Individual principals, teachers and students should not be treated differently due to any component of this measure, or scores on any particular item. Under no circumstances should detailed student response data be made available in its raw form, and no teacher or school professional information should be made public. In addition, caution should be exercised when making high consequence decisions using individual item data, particularly when sample sizes for individual items are small.

7. EVALUATING PRINCIPAL PERFORMANCE: The measure has not been evaluated as a measure of principal performance and should not be used for this purpose.
8. EVALUATING TEACHER PERFORMANCE: The measure does not directly examine teacher performance and should not be used for this purpose.
9. ASSESSING STUDENT PERFORMANCE: The measure does not directly examine student performance and should not be used for this purpose.

Populations Measured. This is an online only student perspectives measure designed for English-speaking student populations in grades 3 through 8. Accommodations may be made for students with disabilities, in order to enable them to complete the online assessments, for

example, students may be allowed to complete the assessment alone or in a quiet place, rather than in classroom settings, or may be given additional time to complete the assessment. Administrators may read the questions and answer choices to students, and may answer questions about the meaning of questions and answer choices, but may not guide student responses. Additional documentation of this issue can be found in the administrator training material provided in Appendix F.

CHAPTER SEVEN:

CALIBRATING THE MEASURE

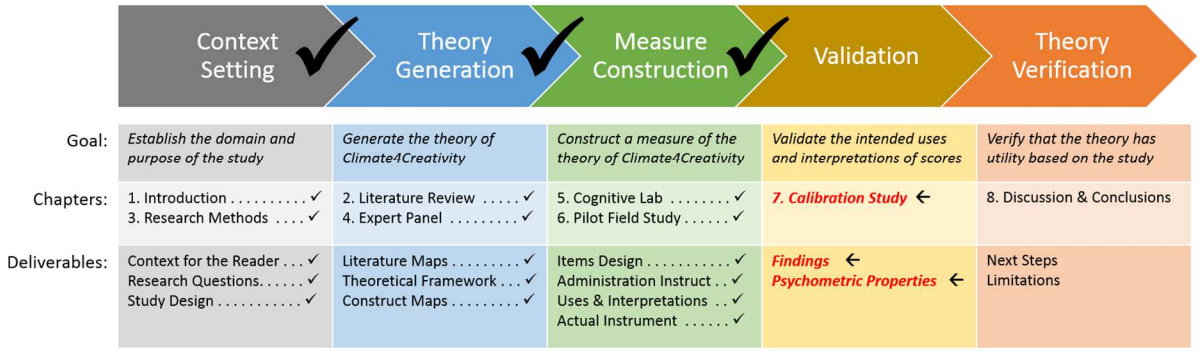


Figure 22. Dissertation Roadmap – Calibration Study.

By enhancing the practice of school improvement with a specific focus on *teaching and learning creativity*, an opportunity emerges to better prepare young people to participate in the social life of the planet, thereby empowering the sustainability of human society. To achieve this, an action-oriented *theory of climate for creativity* was proposed in chapter 4, a measure of the theory was constructed in chapters 5 and 6. The scores from the measure were calibrated in this chapter.

Study IV: Calibration Study

This study was the fourth and final phase of a mixed methods research program designed to explore school climate and the environment for teaching and learning creativity in an elementary school setting. This study was designed to provide quantitative calibration data as input to the following research questions:

RQ1: what are the core elements of healthy school climate & the environment for teaching and learning creativity in a public elementary school setting?

RQ2: how would we measure the degree to which elementary school students believe these core elements exist in their elementary school?

RQ3: what evidence of validity should be collected and what validation arguments should be applied to ensure credibility of such a measure?

RQ4: how might such a measure be employed to actually improve the environment for teaching and learning creativity in a public elementary school?

Introduction. The process used in this study was to test the new Climate4Creativity Student Perspectives measurement instrument, developed as an output from Study I-II of this program of research, with a significant sample of elementary school students. This study was focused on gathering evidence in support of a validation argument for the intended uses and interpretations of the scores yielded from the measurement instruments. This test took the form of a complete major study in twenty-five public elementary schools in Bridgeport, CT.

Participants. A total of 7,915 elementary school students in grades 3-8 attending twenty-five schools in the Bridgeport, CT school system completed the instruments during May and June 2013. The elementary (grades 3-5) instrument was completed by 4,547 students and the middle school (grades 6-8) instrument was completed by 3,360 students. For convenience, several grade K-6 schools used a modified version of the elementary instrument (allowing grade 6 students to complete the simplified elementary school instrument), rather than using both instruments. All schools serving K-7 or K-8 students used both the elementary and middle school instruments. Schools were organized into three administration groups, for ease of training and administration. Tables 11-13 show the number of students completing each instrument at

each school. Demographic profiles of participants by grade level, showing gender, racial identity, Hispanic/Latino status, and home language are provided in Tables 14-17.

Table 11.
Calibration Study Participant Details – Group 1 Schools.

School	**Elementary n_e**	**Middle School n_m**	**Dates**
Barnum School (PK-8)	185	139	May 6-10, 2013
Classical Studies Acad. (K-6)	163	-	May 6-8
Curiale School (K-8)	173	271	May 6-10
Discovery Magnet (PK-8)	132	86	May 6-9
Hall School (K-6)	159	-	May 6-9
Hallen School (PK-6)	157	-	May 6-9
Roosevelt School (PK-8)	157	138*	May 6-10
Bryant School (K-6)	164	-	May 6-10
Group 1 Total	**1,290**	**634**	

Note. * Roosevelt school experienced a duplicate data collection issue, number shown is corrected number.

Table 12.
Calibration Study Participant Details – Group 2 Schools.

School	**Elementary n_e**	**Middle School n_m**	**Dates**
Cesar A. Batalla School (PK-8)	384	349	May 20-Jun 3
Beardsley School (PK-6)	162	-	May 20
Blackham School (PK-8)	282	384	May 20-24
Columbus School (PK-8)	232	129	May 20-30
Dunbar School (K-8)	59	91	May 20-June 7
Luis Munoz Marin (PK-8)	244	300	May 20-30
Group 2 Total	**1,363**	**1,253**	

Table 13.
Calibration Study Participant Details – Group 3 Schools.

School	Elementary n_e	Middle School n_m	Dates
Winthrop School (K-8)	218	228	May 28-31
Black Rock School (K-7)	121	47	May 28-31
Edison School (PK-6)	115	-	May 28-31
Hooker School (K-8)	113	126	May 28-31
Geraldine W. Johnson (PK-8)	239	244	May 28-31
Madison School (K-6)	272	-	May 28-June 3
Read School (PK-8)	241	247	May 28-June 3
Jettie S. Tisdale (PK-8)	168	143	May 28-June 3
Cross School (K-8)	110	139	May 28-31
Multicultural Magnet (K-8)	144	147	May 28-June 12
Park City Magnet (PK-8)	153	152	May 28-June 5
Group 3 Total	**1,894**	**1,473**	

Table 14.
Calibration Study Participant Details – Gender by Grade.

	Grade Level						
Gender	**3**	**4**	**5**	**6**	**7**	**8**	**Total**
Boys	736	722	694	616	637	551	**3,956**
Girls	702	709	697	613	627	559	**3,907**
No Response	14	7	9	6	6	10	**52**
Total	**1,452**	**1,438**	**1,400**	**1,235**	**1,270**	**1,120**	**7,915**

Table 15.
Calibration Study Participant Details – Racial Identity by Grade.

	Grade Level						
Racial Identity*	**3**	**4**	**5**	**6**	**7**	**8**	**Total**
White	547	397	324	242	208	176	**1,894**
Black/African Am.	480	487	448	413	468	453	**2,749**
Native American	62	51	63	43	58	45	**322**
Asian American	42	36	44	52	59	27	**260**
Pacific Islander	7	18	12	12	22	18	**89**
Other Race	304	450	529	550	528	466	**2,827**

Note. * Respondents could select multiple racial identities.

Table 16.
Calibration Study Participant Details – Hispanic/Latino Status by Grade.

Hispanic Latino	**Grade Level**						
Status	**3**	**4**	**5**	**6**	**7**	**8**	**Total**
Hispanic Latino	609	682	703	642	659	597	**3,892**
Not Hispanic Latino	806	708	655	574	589	511	**3,843**
No Response	37	48	42	19	22	12	**180**
Total	**1,452**	**1,438**	**1,400**	**1,235**	**1,270**	**1,120**	**7,915**

Table 17.
Calibration Study Participant Details – Home Language by Grade.

Home Language	**Grade Level**						
English	**3**	**4**	**5**	**6**	**7**	**8**	**Total**
English at Home	1,077	1,048	1,056	929	970	881	**5,961**
Not English at Home	345	362	323	287	274	227	**1,818**
No Response	30	28	21	19	26	12	**136**
Total	**1,452**	**1,438**	**1,400**	**1,235**	**1,270**	**1,120**	**7,915**

Procedures and Materials. Written approval to work in the Bridgeport schools was provided by the superintendent and board of education. The principals of the twenty-seven elementary schools in the Bridgeport, CT school district were then contacted by the district Positive Behavior Interventions and Supports (PBIS) team leader, and all but two provided site approval for the study to be held on school premises during the regular school day. Each of the twenty-five participating principals provided information to be used for customization of the instruments for their school, as well as the name or names of their onsite coordinator, who would manage the administration of the instrument for their school. A three hour training session was held with all onsite coordinators, and a number of principals in attendance, and a detailed coordinator administration packet was provided to each onsite coordinator. A secure website was provided for the 40 onsite coordinators and principals to provide them with easy access to

materials, links to electronic instruments, daily completion statistics, and other useful administrator resources.

Onsite coordinators worked within their schools, supported by the district PBIS team leader, to schedule students to complete the instruments during their normal school day. I was never required to be on site during administration of the instrument, but checked in periodically with onsite coordinators via phone and email. Typical school completion involved all students from a common class reporting to the school computer lab during their regular scheduled class period and completing the instrument in the lab along with their classmates. Clear instructions were read to students by the onsite coordinator. Students were informed that they did not have to complete the survey and that their answers would be kept confidential. In addition, individual informed assent was obtained from students on the first page of the electronic survey. On completion of the survey, students returned to their normal scheduled class, and the onsite coordinator prepared the computers for the next scheduled class of students.

Several schools did not, at the time, have dedicated computer labs. In these schools, students cycled through a survey location, sometimes the school library, or a reading resource room, and completed the survey in smaller groups, based on the availability of computers in the school. In one case, students completed the survey using the school's iPads, sitting in an area near the faculty lounge. Onsite coordinators worked to maintain a quiet, private space for students, as described in their administration instructions.

If students were absent on the day of administration for their class, they were given an opportunity to complete the survey at another time, where possible. I monitored completion statistics on a daily basis, and checked completion rates by school and grade level. In circumstances where a particular grade level seemed over or under represented, I conducted a

brief investigation to ensure that no serious issues had occurred during administration of the measure. In one case, with 4th grade students at Roosevelt school, a national scope issue with the SurveyMonkey tool occurred during administration of the instrument. A careful investigation of the data was completed to ensure no duplicate data was included in the completed analysis. The complete instrument (v1.2) used for the calibration study is provided in Appendix E. Informed consent and site approval materials are provided in Appendix G.

Data Collection. Students completed an online instrument, version 1.2. The instrument was delivered to students using SurveyMonkey online survey tool. Students in grades 3-5 completed the elementary instrument (C4C/SPE), and students in grades 6-8 completed the middle school instrument (C4C/SPM), with the exception of grade 6 students in schools serving only grades K-6 and PK-6: These students completed the elementary instrument (C4C/SPE).

Data for each school/instrument combination was extracted from SurveyMonkey into an Excel (Microsoft, 2013a) spreadsheet for each school/instrument combination, using condensed numerical representations of the full data. These Excel files were individually manipulated to remove all respondent identification data, and to relabel data according to the specification given in Appendix E. A two digit code was added to the beginning of each SurveyMonkey Respondent ID in the Excel file to identify the school attended, while maintaining uniqueness of each respondent. Data were then consolidated into a single Winsteps readable text file for each instrument, each file containing data from multiple schools. Five Winsteps control files were then created for each instrument, one for each construct within each instrument and one for the overall Climate4Creativity measure. No attempt was made to combine elementary and middle school data into a single measure for a combination school. Reliability, effective distribution of items and range of the measures overall, category function for all items, and fit characteristics of

items within each construct were analyzed using Winsteps to ensure the quality of the measurement scores extracted from the instrument.

Data Analysis. Winsteps was run again for each control file in turn, producing a complete Winsteps person file output of each instrument/construct combination. These ten output files were then consolidated into a single output file for each instrument, by combining data using the Respondent ID as a common key across data files. This created a single file for each instrument, keyed by school and respondent id, and containing matched sets of the five measures for each respondent, along with five sets of person fit and data quality characteristics. Missing response data was removed from the data files, and an analysis of extreme high and extreme low responses was conducted. Respondents who specified category 5 or category 1 in response to all items were removed from the file. Students who responded all 5's or all 1's to an individual construct, but did not respond all 5's or all 1's to other constructs were included in the data. This data cleansing process was deliberately conservative in what items were removed, so as to preserve as much respondent information as feasible, within the boundaries of data quality. Table 18 documents the data cleansing process used in preparation for this analysis.

Table 18.
Data Cleansing Results.

	Middle School		**Elementary**		**Totals**	
Initial Data Count	3,263		4,684		**7,947**	
- Missing Data:	-141	(4.3%)	-154	(3.7%)	**-295**	**(3.7%)**
- Extreme Highs:	-2	(0.1%)	-2	(0.0%)	**-4**	**(0.1%)**
- Extreme Lows:	-1	(0.0%)	0	(0.0%)	**-1**	**(0.0%)**
Clean Data Count	**3,119**		**4,528**		**7,647**	

Once the data were clean, four types of analysis were conducted. First, reliability and other psychometric properties were analyzed. Second, an analysis of the strength of relationships between constructs at the individual student level was completed. Third, a school level analysis was conducted to evaluate the strength of relationships between constructs at the school level. Finally, a series of comparisons were conducted at the school level between publically available discipline data and prior school climate assessments, to evaluate the consistency of the new instruments in use in the school system.

Reliability. The reliability of an instrument is a statistic designed to estimate the true consistency of the instrument in different uses under different circumstances. A reliability coefficient value of 1.0 implies that the instrument is perfectly reliable each time it is used. The conventional reliability coefficient used in classical test theory is Cronbach's Alpha reliability. Rasch measurement modeling provides an alternative error-adjusted person separation reliability, which provides a lower bound estimate of the reliability, since it tends to underestimate true reliability, whereas Cronbach's Alpha tends to overestimate true reliability. These two reliability coefficients can be used as boundary limits for the reliability of a measure (Linacre, 2012).

Item Polarity & Item Fit. Item polarity measures the degree to which individual items within a measure move in the same direction as the measure itself. Item fit analysis examines the degree to which individual items fit the assumptions of the Rasch measurement model. Standard errors were analyzed using Standard Errors of Measurement (SEM) curves, and an analysis of raw score to measure ogives was used to examine the ability of the measure to differentiate between measures. This second analysis was designed to ensure that the standard errors of measurement did not overwhelm the measures themselves, invalidating the logit scales.

Category Functioning. Category functioning analysis examines the extent to which the response categories for each item in a measure effectively discriminate between subcategories of respondents. There should be clear and well-ordered thresholds between categories (known as Andrich Thresholds), and categories should function in the right order (Bond & Fox, 2007).

Measure Range. The range of effectiveness of the measure is examined by reviewing the bottom and top halves of the respondents to the measure, and mapping the locations of the most and least difficult items on the logit scale. The effective range of the measure is seen as the range from the least difficult item for the bottom half of the respondents to the most difficult item for the top half.

Differential Item Functioning (DIF) Analysis. This analysis examines the fairness of item responses, with particular focus on whether respondents in different subclasses respond differently to the measure. DIF analysis was completed for the two overall measures using gender and ethnic groups as differentiators.

Correlation Analysis by Construct for Individuals. Once the psychometric quality checks were completed, a scatterplot was created for each pair of constructs, and a correlation analysis, using Pearson's *r* was conducted to examine the relationships between matched pairs of construct measures for the entire school system, at the individual student level. Pearson's *r* effect sizes were computed for each correlation, and a corresponding coefficient of determination, r^2, was computed for each combination of constructs. This combination of statistics was used to examine the strength of the hypothesized relationships between each construct, and the proportion of variance shared by the two variables. These correlational analyses were intended to investigate the relationships between constructs at the individual level, across all students in all schools in the study.

Correlation Analysis by Construct for Schools. Mean scores were then computed by school/instrument and a series of boxplots were created for each school/instrument. A correlation analysis was conducted on a school/instrument basis between mean scores for each combination of constructs, in an attempt to understand the strength of schoolwide relationships between safety, community, learning environment, and the environment for teaching and learning creativity. Pearson's *r* effect sizes were computed for each correlation, and a corresponding coefficient of determination, r^2, was computed for each combination of constructs. This combination of statistics was used to examine the strength of the hypothesized relationships between each construct, and the proportion of variance shared by the two variables. The school/instrument correlational analysis was intended to investigate the schoolwide relationships between constructs, by comparing the data for each school, rather than at an individual level.

Reference Checks. The school level analysis was rounded out by matching schoolwide student perspective measures (safety and community measures) with publically available data on school discipline, and by matching overall Climate4Creativity results with results from the Comer School Climate Instruments used in some of the schools two years before this study.

Psychometric Properties of the Instruments. Table 19 shows the reliability coefficients for each instrument and construct. Note that where I refer to an item or an answer choice in this section, I will show the item code in plain text and the item text in italics. A full definition of the items and answer choices can be found in Appendix E.

Table 19.
Reliability Results for C4C Middle School & Elementary Instruments.

Instrument	Construct	Lower Bound: Person separation reliability, r	Upper Bound: Cronbach's Alpha
Middle School	Overall	0.95	0.98
	Safety	0.91	0.96
	Community	0.85	0.90
	Learning	0.86	0.93
	Creativity	0.83	0.93
Elementary School	Overall	0.91	0.97
	Safety	0.87	0.96
	Community	0.76	0.91
	Learning	0.56	0.89
	Creativity	0.69	0.90

Due to the relatively low person separation reliability of the learning (r = 0.56) and creativity (r = 0.69) constructs within the Elementary School instrument, a detailed analysis of these components of the measure was completed. The fit characteristics were acceptable in both cases, but detailed examination of the Andrich Thresholds showed poor category performance on several of the items, implying that a significant proportion of elementary school students were unable to distinguish between some of the response categories. For example, the learning construct showed misordered Andrich Thresholds at the lower end of the scale and very close threshold values at the upper end of the scale, as shown in Table 20.

Table 20.
Andrich Threshold Analysis of the Learning Construct on C4C/SPE.

Category Structure	Category Threshold	Empirical Andrich Threshold
1-2-3-4-5	1-2	- 0.48
	2-3	- 0.73
	3-4	+0.59
	4-5	+0.63

Clearly, if the category structure were operating properly as a 5-point scale, the Andrich Threshold for category 1-2 would be lower than the Andrich Threshold for category 2-3, and it would be easy to distinguish between the thresholds for categories 3-4 and 4-5. In order to address this issue, the category structure for the learning and creativity constructs were examined using Winsteps, until an acceptable category structure was observed. The process of category restructure is illustrated in Table 21.

Table 21.
Category Structure Analysis for the Learning Construct of C4C/SPE.

Category Structure	**Person separation Reliability, r**	**Category Threshold**	**Empirical Andrich Threshold**
1-2-3-4-5	0.56	1-2	- 0.48[1]
		2-3	- 0.73[1]
		3-4	+0.59[2]
		4-5	+0.63[2]
1-2-2-3-4	0.61	1-2	- 1.68
		2-3	+0.88[1]
		3-4	+0.80[1]
1-2-2-3-3	0.35[3]	1-2	- 1.18
		2-3	+1.18
1-2-2-2-3	0.65	1-2	- 2.02
		2-3	+2.02

Notes. [1]Misordered thresholds. [2]Thresholds too close. [3]Person separation reliability too low.

Based on this analysis, the category structures for both the learning and creativity constructs were adjusted to use a 1-2-2-2-3 structure, essentially merging categories 2, 3, and 4 into a single central category and maintaining categories 1 and 5 as lower and upper categories. In both cases this led to strong category functioning for these constructs and to improved person separation reliability, r. The adjusted learning construct now exhibited person separation reliability of r = 0.65 and the adjusted creativity construct exhibited person separation reliability

of r = 0.72. Based on this adjustments, reliability results for the elementary instrument are shown in Table 22.

Table 22.
Reliability Results for the Adjusted C4C/SPE Instrument.

Instrument	Construct	Lower Bound: Person separation reliability, r	Upper Bound: Cronbach's Alpha
Elementary School	Overall	0.91	0.97
	Safety	0.87	0.96
	Community	0.76	0.91
	Learning[1]	0.65	0.96
	Creativity[1]	0.72	0.93

Note. [1]Construct category structure adjusted to 3-point scale.

Despite this change, the learning and creativity constructs on the elementary school measure still exhibited lower than anticipated person separation reliability levels, although their Cronbach's Alpha reliability scores were excellent. This is in contrast to the learning and creativity construct measures on the middle school instrument – the only significant difference being the number of items. The middle school instrument asks respondents the learning and creativity items twice, once in reference to their favorite subject and once in reference to their least favorite subject, and has one additional item related to their anticipated grade level in each class. The elementary learning and creativity items designs have 9 items each, the middle school learning design has 20 items, and the middle school creativity design has 18 items.

As an experiment, the elementary learning and creativity constructs were combined into a single structure with 18 items, to examine if this would improve the psychometric properties of the measure overall, without significant impact on the intended uses and interpretations of the instrument. This new merged learning and creativity construct exhibited significantly better reliability (r = 0.82, $\alpha = 0.97$), and had just a single misfitting (infit = 1.84, outfit = 1.97) item,

23D-TIME: *The work I do in class makes me lose track of time*. By eliminating this single item from the combined measure, the new merged learning and creativity construct exhibited excellent fit and reliability as a new construct. This item was removed from the overall measure, improving the reliability of the overall measure, without significant loss of meaning. Table 23 shows reliability scores based on merging the learning and creativity constructs into a single construct.

Table 23.
Reliability Results for the Final Adjusted C4C/SPE Instrument.

Instrument	Construct	Lower Bound: Person separation reliability, r	Upper Bound: Cronbach's Alpha
Elementary School	Overall	0.92	0.97
	Safety	0.87	0.96
	Community	0.76	0.91
	Learning & Creativity[12]	0.82	0.97

Note. [1]Construct category structure adjusted to 3-point scale. [2]Constructs merged.

All nine of the constructs were examined for fit, and found to have good fit characteristics, with no item infit or outfit MNSQ values in excess of 1.70 or below 0.50 on any construct. No item polarity issues were identified on any of the constructs. Table 24 shows the fit characteristics for the elementary school overall measure, and Figure 23 shows the corresponding empirical item-person map. Note, empirical fit characteristics and empirical item-person maps for all measures are provided in Appendix H and Appendix I.

Table 24.
Fit Characteristics for the Highest and Lowest Fit Items; Elementary School Overall Measure.

Items	Meas. (logits)	Error	Infit MNSQ[1]	Z(Std)[2]	Outfit MNSQ[1]	Z(Std)[2]	Corr.
6C-WKHD: *My family members encourage me to work hard*	-0.92	0.02	1.63	9.90	1.62	9.90	0.25
7C-STU: *Students in other grades help me*	1.24	0.01	1.35	9.90	1.52	9.90	0.32
17C-PUN: *Adults punish the bully*	-0.23	0.02	1.41	9.90	1.52	9.90	0.39
6D-TTCH: *My family members talk to my teachers about me*	0.37	0.01	1.30	9.90	1.42	9.90	0.27
14-TALK: *How often do you hear students talk about being mean to others?*	1.11	0.01	1.15	7.70	1.42	9.90	0.29
24-CRE: *How important do you think it is to be creative in class?*	0.23	0.01	1.15	7.40	1.42	9.90	0.26
7A-FRND: *My closest school friends help me*	0.25	0.01	1.27	9.90	1.38	9.90	0.29
7E-ADLT: *Other adults in my school help me*	0.47	0.01	1.32	9.90	1.38	9.90	0.36
6E-EVNT: *My family members come to school events*	0.72	0.01	1.21	9.90	1.33	9.90	0.30
17D-TALK: *Adults talk to the class about bullying*	0.02	0.01	1.26	9.90	1.31	9.90	0.44
--- BETTER FITTING ITEMS OMITTED ---							
11A-CLS: *How safe do you usually feel in your classroom?*	-0.63	0.02	0.88	4.50	0.80	6.40	0.48
11D-PLGR: *How safe do you usually feel outside in the playground?*	0.10	0.01	0.88	6.10	0.86	6.40	0.48
8F-RSST: *Students are respectful to teachers*	0.21	0.01	0.83	9.70	0.85	7.00	0.47
8E-KNST: *Students are kind to teachers*	0.22	0.01	0.79	9.90	0.83	8.40	0.47
11B-HALL: *How safe do you usually feel walking in the hallways?*	-0.03	0.01	0.79	9.90	0.79	9.20	0.48
19-SAFE: *How safe do you feel overall at your school?*	0.03	0.01	0.79	9.90	0.77	9.90	0.59
10B-B4: *How safe do you usually feel before classes start for the day?*	-0.33	0.02	0.77	9.90	0.77	8.60	0.45
23A-FUN: *The work I do in class is fun*	0.29	0.01	0.74	9.90	0.75	9.90	0.55
8B-RSSS: *Students are respectful to other students*	0.57	0.01	0.69	9.90	0.74	9.90	0.50
8A-KNSS: *Students are kind to other students*	0.53	0.01	0.69	9.90	0.73	9.90	0.52
Mean	*0.00*	*0.02*	*1.03*	*0.80*	*1.03*	*0.00*	-
Standard Deviation	*0.43*	*0.00*	*0.18*	*6.60*	*0.22*	*6.50*	-

Notes. [1]Mean square fit characteristic. [2]Z scores are standardized.

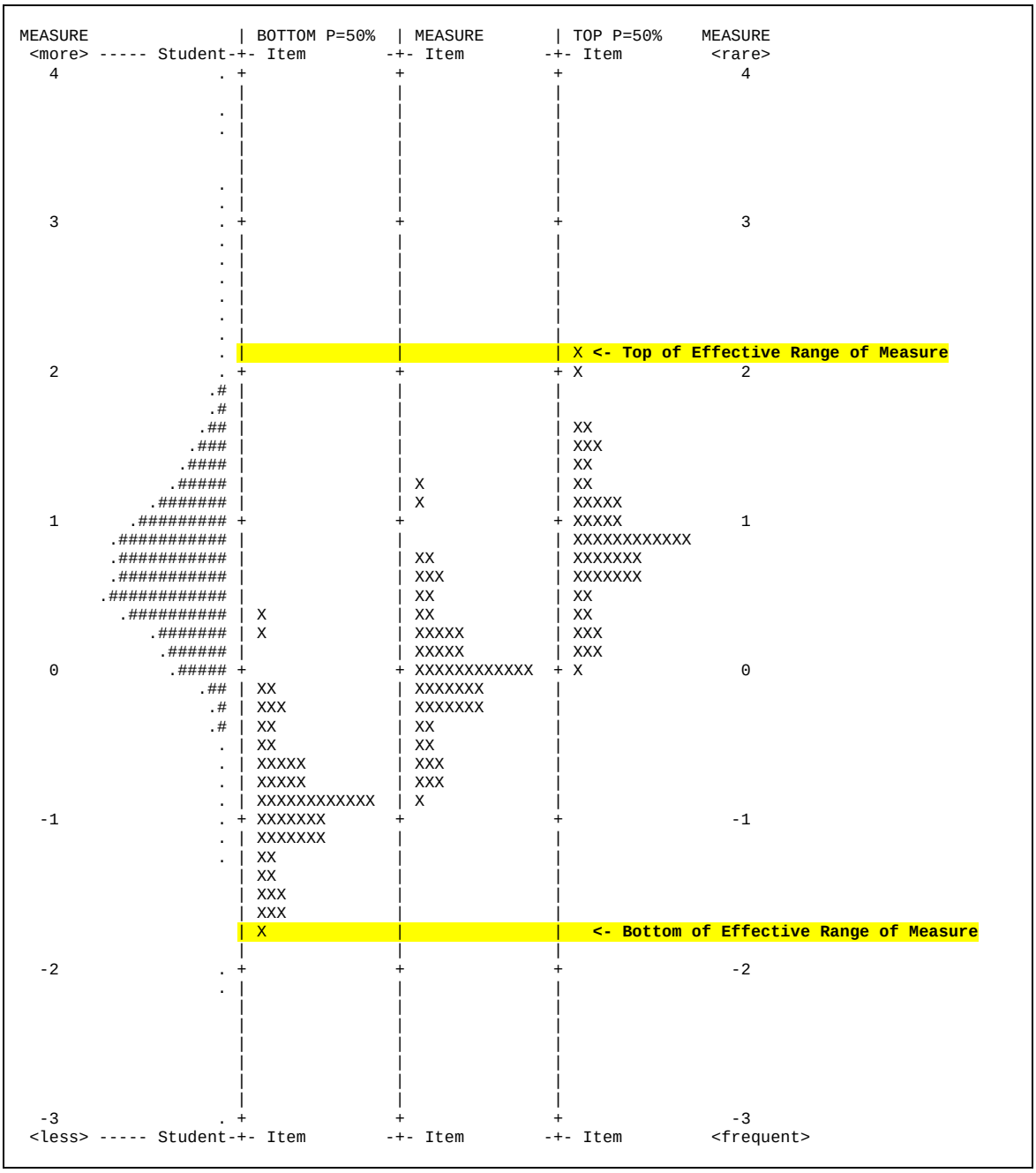


Figure 23. Empirical Item-person Map of Elementary School Overall Measure.

Note. Each "#" represents 38 students. Each "." represents 1 to 37 students. Item wording can be found in Appendix E. Students at the top of the map have a high feeling of climate for creativity overall (ability), students at the bottom have a low feeling of climate for creativity overall.

Standard errors were analyzed using a Standard Error of Measurement (SEM) curve.

Under normal functioning conditions, the standardized error of measurement should not

overwhelm the estimated measures, although it is clear that at the highest and lowest levels of a measure, we expect to see a higher level of standard error, since, on a well-fitting scale, less respondents tend to pick very high or very low scores. This is illustrated in Figure 24 for the middle school safety measure, which exhibits the expected U-shaped SEM curve. SEM curves for all measures are provided in Appendix J.

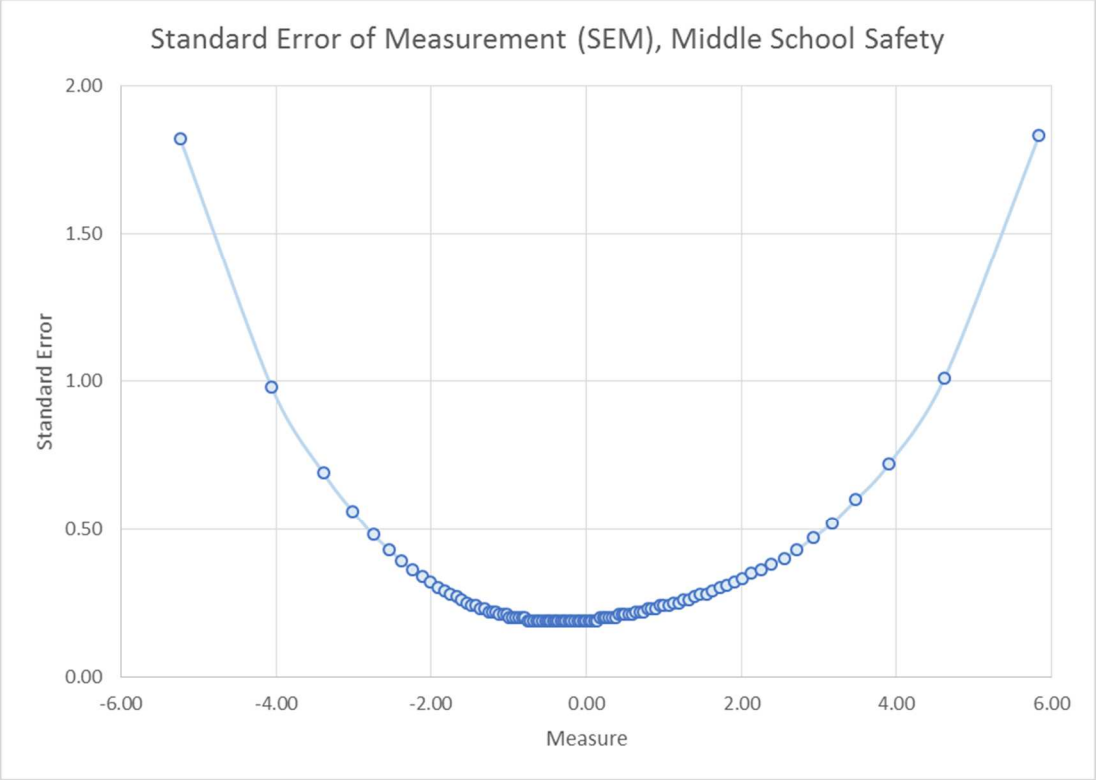


Figure 24. Standard Error of Measurement (SEM) Curve.

A subsequent analysis, comparing raw scores to estimated measures for the same construct measure, shows a classic S-shaped ogive, providing further evidence of proper scale function. The generally linear middle area of the measurement scale is as expected, and typical ceiling and floor effects are illustrated by the increased curvature at either end of the scale. Figure 25 illustrates this analysis for the middle school safety measure.

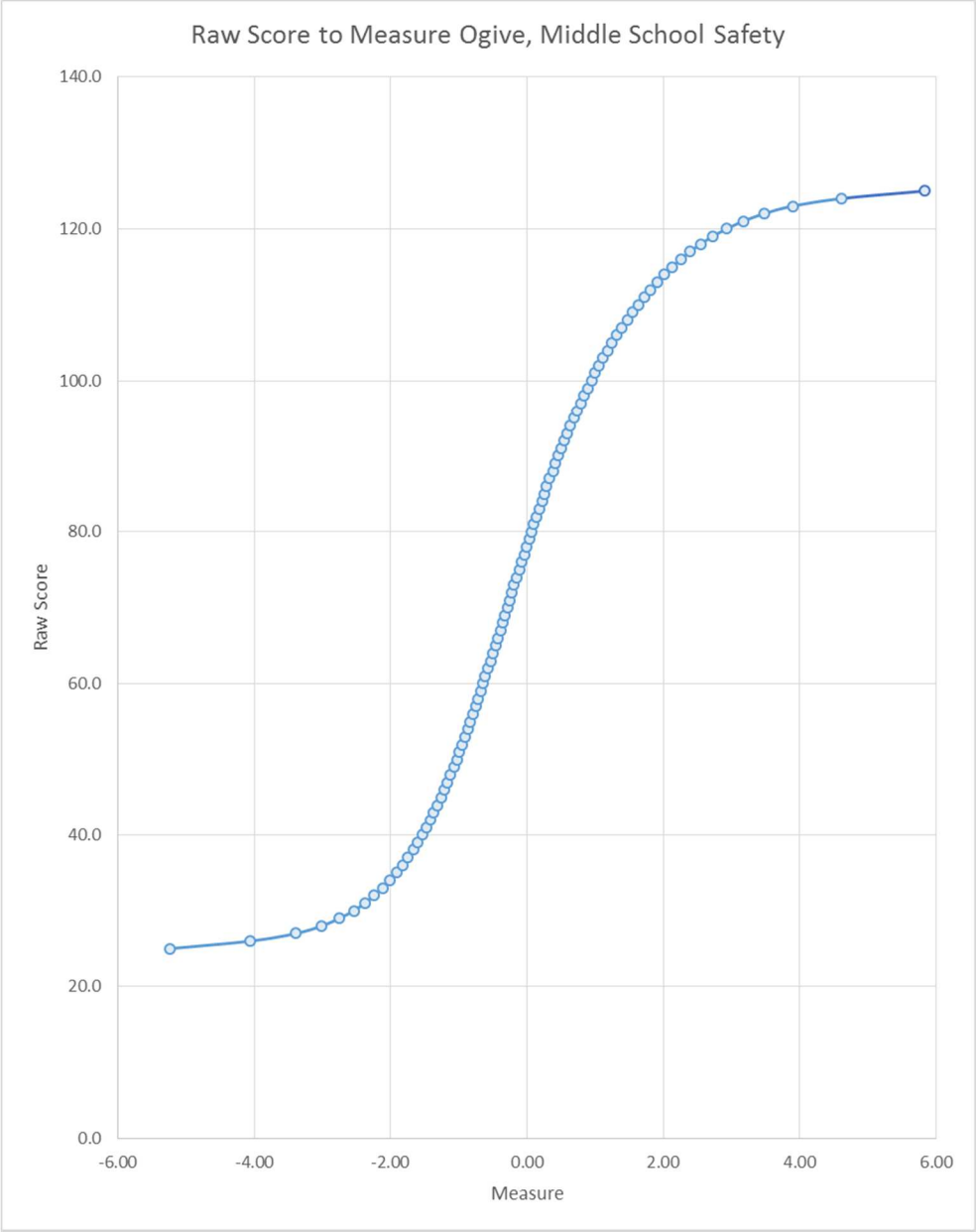


Figure 25. Raw Score to Measure Ogive.

By combining the information in the SEM curve and the raw score to measure ogive, an analysis of the three core data points – raw scores, measures, and standard error – reveals the functioning of the measure as an interval scale, and the degree to which the noise generated by the standard error is contained in the extremes of the measure. This provides compelling

evidence of the efficacy of the measures. Figure 26 illustrates this multipoint analysis using a bubble ogive, with bubble diameter representing the standard error, overlaid on the raw score to measure ogive.

Figure 26. Raw Score to Measure Ogive, Showing Standard Error.

Note. Bubble diameter indicates the size of the standard error of measurement.

No category functioning issues were identified on any of the items, once the new elementary school construct and category structure was implemented in full. Winsteps analysis of the distribution of students in both instruments, showed the items to be fairly evenly distributed across the range of the measure. Items were slightly easy for the middle school respondents in the study, implying that some small number (< 2%) of students might want to show a higher level of the overall measure than is reportable, since the effective maximum range

of the items leads to slight compression at the top of the scale. The measure overall was well matched with the elementary school respondents. The bottom of the scale range in both instruments appears appropriate for the distribution of students found in the study. Table 25 illustrates this analysis for the distribution of students across the middle schools measures.

Table 25.
Middle School Instrument Student Distribution

	Left Outliers	**Left 1-2σ**	**Mean ± 1σ**	**Right 1-2σ**	**Right Outliers**
Overall	44 (1.4%)	253 (7.9%)	2,611 (81.3%)	255 (7.9%)	58 (1.7%)
Safety	42 (1.3%)	236 (7.4%)	2,546 (79.4%)	312 (9.7%)	69 (2.2%)
Community	50 (1.6%)	258 (8.0%)	2,625 (81.5%)	221 (6.9%)	65 (2.0%)
Learning	30 (1%)	193 (6.2%)	2,584 (82.5%)	241 (7.7%)	84 (2.7%)
Creativity	39 (1.3%)	209 (6.7%)	2,573 (82.3%	201 (6.4%)	104 (3.3%)

Figure 27 shows an example Item-person Map, in this case for the middle school instrument overall measure. This form of the item-person map shows the fit of the top half, bottom half, and whole respondent pool on the same map, providing evidence of the range of functioning of the instrument. It is clear that the instrument is effective for measuring the attitudes of the vast majority of respondents in the population. Item-person maps are provided for all measures in Appendix I.

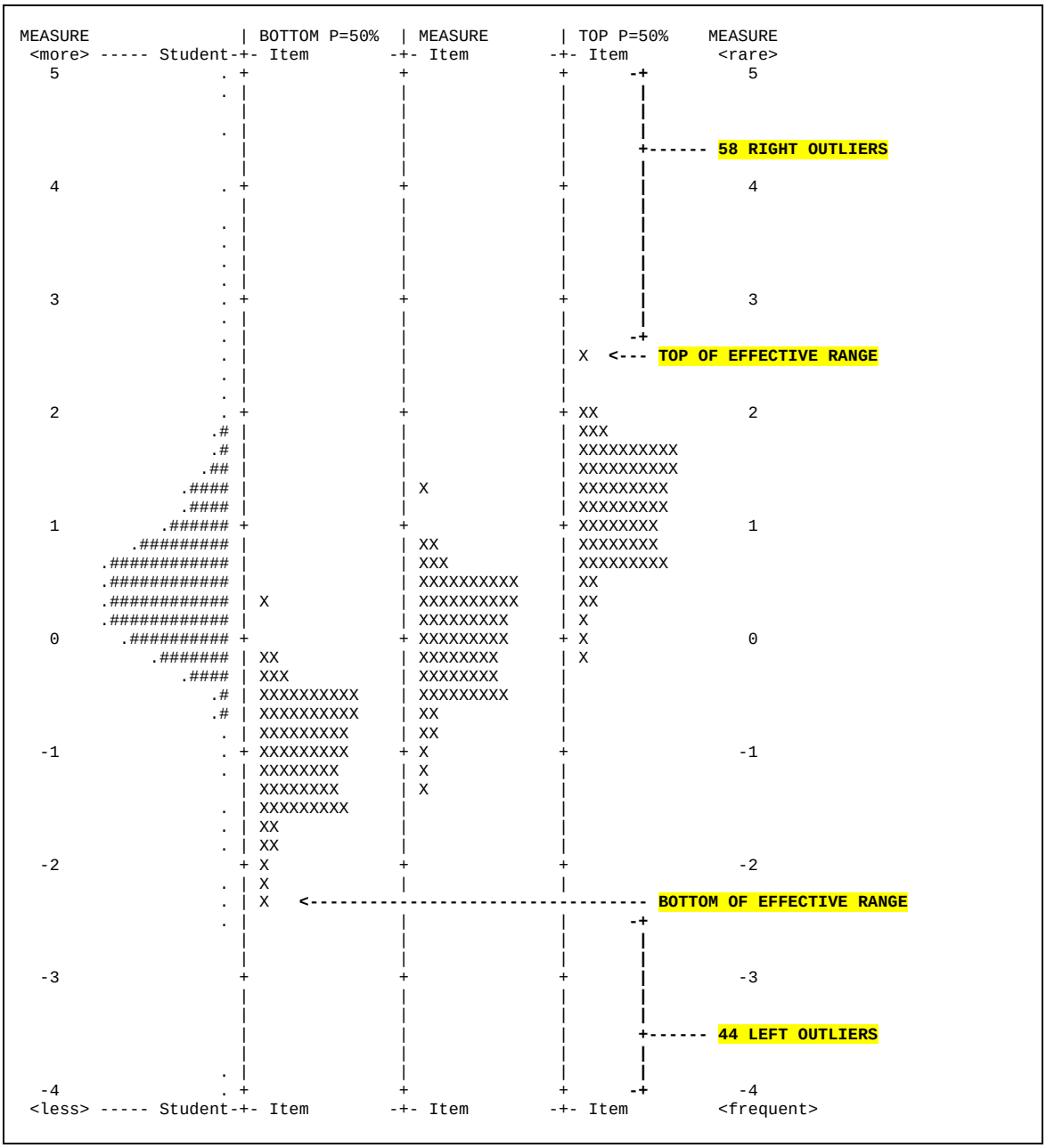


Figure 27. Empirical Item-person Map of Middle School Overall Measure.

Note. Each "#" in the Student column represents 29 students. Each "." represents 1 to 28 students.

Differential Item Functioning (DIF) was examined using Winsteps for the overall measures in both instruments. Since all the items for the four constructs were included in the overall measures, it was determined that it was not necessary to repeat this analysis for the other

constructs. This analysis was completed for gender using a simple male/female contrast, and for ethnicity using a compound white/non-white contrast. The Mantel (1963) procedure was employed, providing a log-odds estimator of DIF size and significance between groups based on a cross-tab of the observations from the two groups of respondents. Zieky (1993) recommended interpretations of DIF levels in a Rasch measurement context were used: A logit difference value below 0.426 is called *negligible*, values between 0.426 and 0.638 are called *intermediate*, and values above 0.638 are called *large*. A DIF Size chart was used for each construct to examine the differences between the average measures across all respondents and the DIF groups being compared. Figure 28 shows the DIF size chart for the elementary school overall construct, male/female contrast.

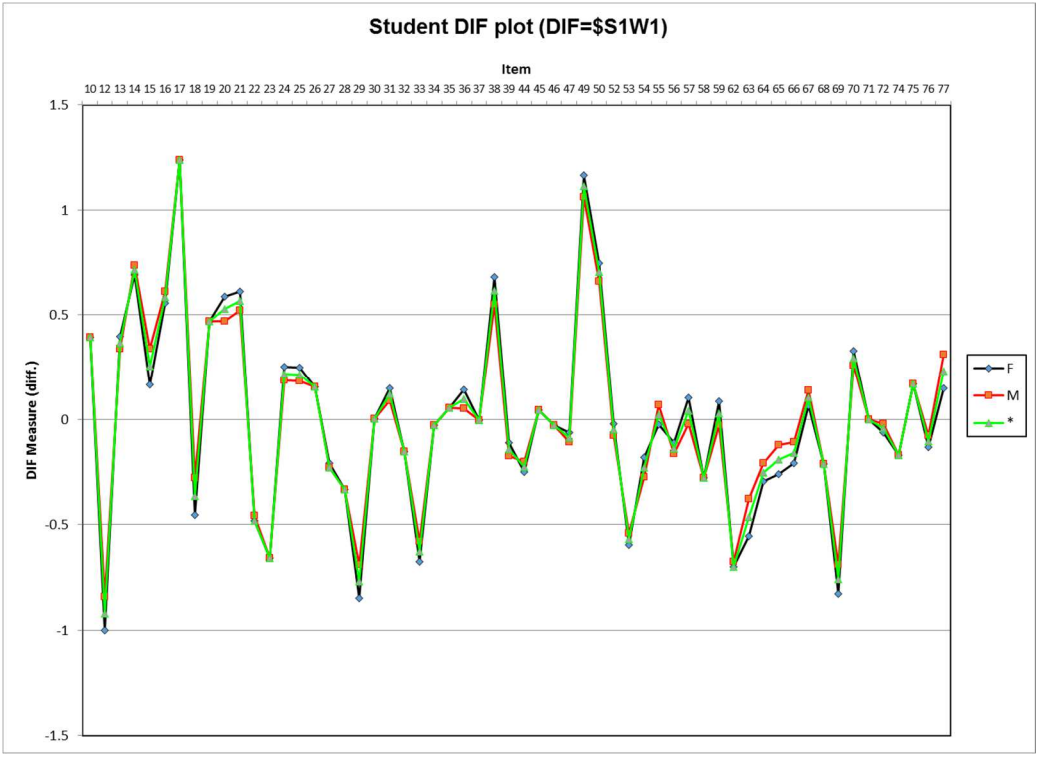


Figure 28. DIF Size Chart for Elementary Overall Measure, Male/Female Contrast.

Notes. F – Female respondents; M – Male respondents; * - Average for all respondents.

It is clear from the diagram that the differential sizes for this instrument are very small, although there are some minor differences. The largest difference noted in this analysis is between boys and girls, where girls, on average, were more likely to say their friends helped them with their homework. This was a DIF size of 0.34, classified as negligible. The largest DIF value in the White/Non-White contrast was on the item asking if family members encouraged the respondent to work hard. Non-White respondents were less likely to score highly on this item, but with a negligible DIF size of 0.22 (middle school) and 0.20 (elementary). This analysis demonstrates that the items operate in an unbiased manner with respect to gender and white/non-white ethnicity.

Correlation Analysis by Construct at the Individual Level. This analysis examined the effect sizes of the relationships between the constructs within each instrument, so as to establish the strength of the relationships between each measure. This analysis provides direct input to the theory of Climate4Creativity, which predicts the relative strength of those relationships. First, each construct was compared with the overall measure within the instruments, so as to demonstrate that the overall Climate4Creativity measure was strongly related on the underlying constructs of the instrument. This analysis utilized both the Pearson Correlation Coefficient, r, and the Coefficient of Determination, r^2 (r-squared). Pearson's r was used as a primary measure of the effect size of the relationship between the constructs and the overall measure, and r-squared was used to describe the degree to which the measures share variation in the empirical data. Cohen (1988, 1992) suggests that when using Pearson's r as an effect size, values above 0.5 should be treated as large effect, and below 0.3 as small effect. This analysis therefore shows large effect sizes for all components of the model, implying a high

likelihood that the core constructs do, in fact, form the basis of the overall measure of Climate4Creativity in a school. Results of this analysis are shown in Table 26.

Table 26.
Overall Measure to Core Construct Cross-correlations.

Instrument	Measure	Correlation to Overall Measure, r^a	Coefficient of Determination, r^2
Middle School	Safety	0.81	0.66
	Community	0.77	0.59
	Learning	0.82	0.67
	Creativity	0.80	0.64
Elementary	Safety	0.87	0.76
	Community	0.76	0.58
	Learning & Creativity	0.75	0.56

Notes. [a]Pearson Correlation Coefficient.

Next, the cross-construct relationships were examined, first with scatterplot charts, and subsequently with correlation calculations. Using both the Pearson's r and the r-squared coefficient of determination, a large effect size ($r = 0.84$, $r^2 = 0.71$) was observed between the learning and creativity measures in the middle school instrument, illustrated by the scatterplot shown in Figure 29.

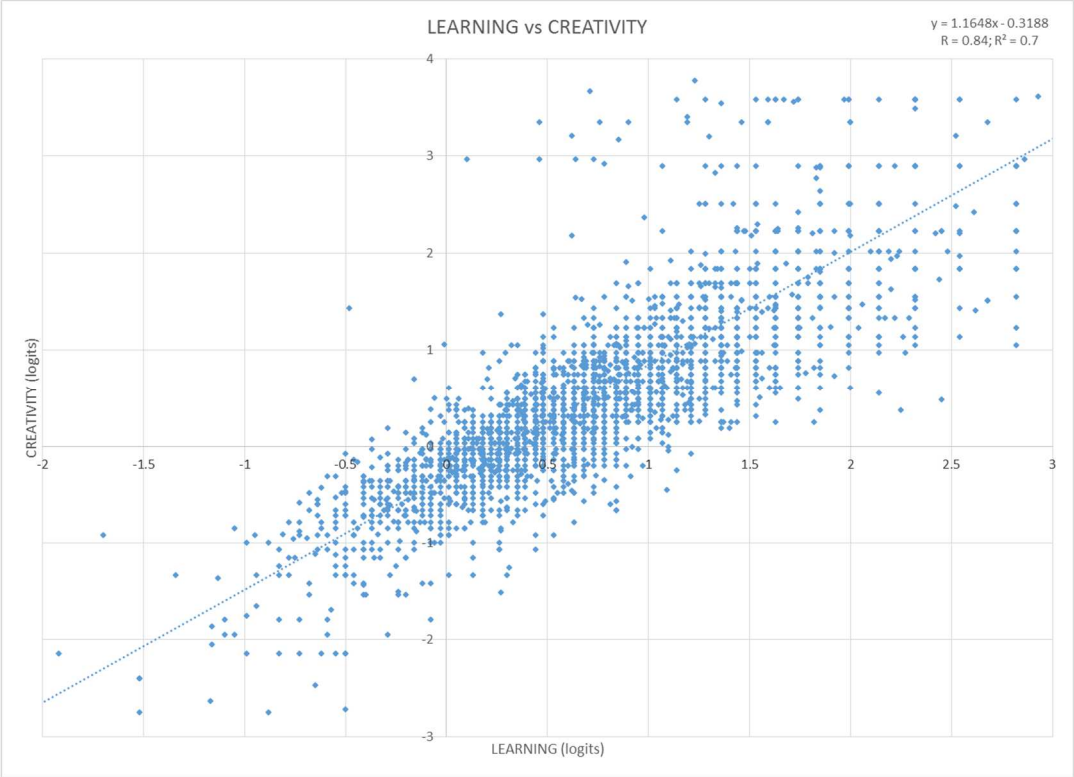


Figure 29. Scatterplot of Learning vs Creativity on the Middle School Instrument.

A second large effect size ($r = 0.61$, $r^2 = 0.37$) was found between the safety and community measures in the middle school instrument, and moderate effect sizes ($0.48 \leq r \leq 0.49$), were found between the safety/community and learning/creativity measures. These weaker relationships are illustrated by the relationship between the community and learning constructs, shown in the scatterplot in Figure 30.

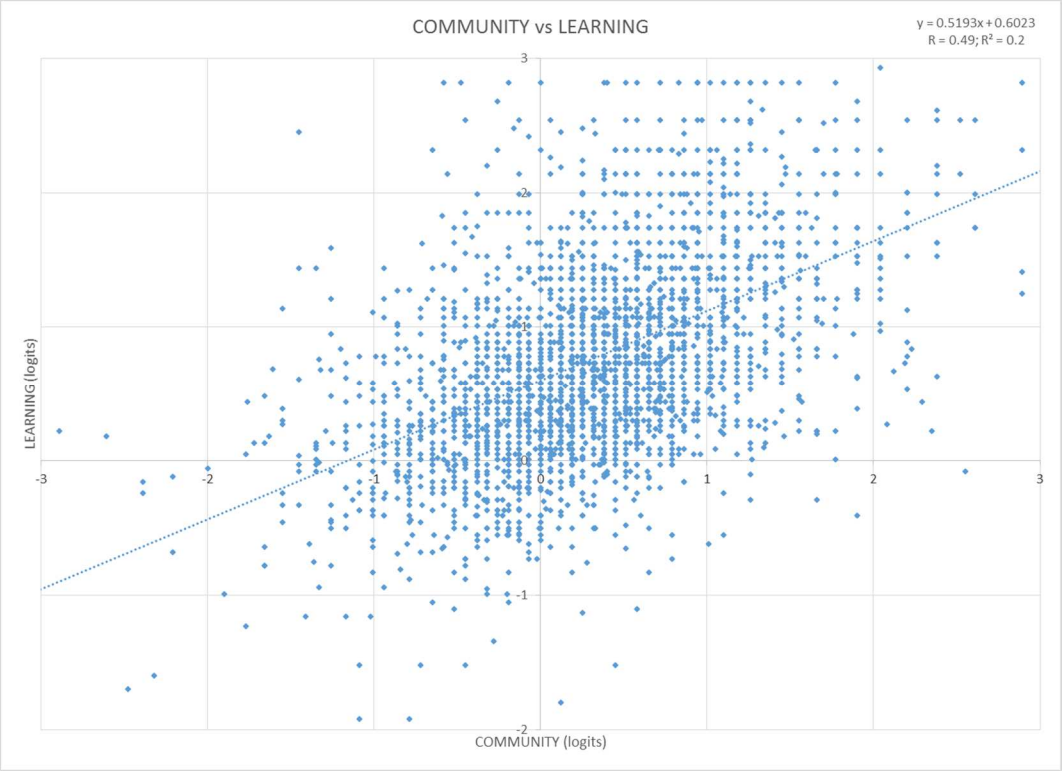


Figure 30. Scatterplot of Community vs Learning on the Middle School Instrument.

Similar moderate effect sizes ($0.51 \leq r \leq 0.53$) were observed between all three constructs in the elementary school instrument. The results of this cross-correlation analysis are shown in the cross-correlations table, Table 27.

Table 27.
Construct Cross-correlation Analysis Findings, Individual Level.

		Effect Sizes: Pearson's r (r-squared, r^2)		
Instrument	**Construct**	Community	Learning	Creativity
Middle School	Safety	0.61 (0.37)	0.49 (0.24)	0.48 (0.23)
	Community	-	0.49 (0.24)	0.48 (0.23)
	Learning	-	-	0.84 (0.71)
Elementary	Safety	0.51 (0.26)	0.52 (0.27)[1]	
	Community	-	0.53 (0.28)[1]	

Note. [1]Combined Learning & Creativity Construct used in Elementary Instrument.

Correlation Analysis by Construct at the School Level. Another way to examine the findings from the Climate4Creativity instruments is to interpret them as measures of schoolwide safety, community, learning, and creativity. This analysis was completed by analyzing each school individually, and comparing the mean responses and distribution of responses for each school. This analysis showed that schools exist on a continuum of each construct, with some schools scoring better than others.

One effective way to present such complicated findings is the use of comparative boxplots by school/instrument and construct. This allows us to graphically assess the degree of difference between schools. Boxplots are useful because they show a large quantity of information in a relatively easy to absorb way. The width of the boxplot shows the distribution of responses for a school, a wider distribution shows a higher degree of variability in student responses, and a narrower distribution can be interpreted as more agreement about the measure. Higher median values (the line in the middle of the box section of the boxplot), shows the middle score within the school, and the mean (the diamond in the center of the boxplot) indicates the average score within the school. A comparative boxplot of middle schools, on the overall Climate4Creativity measure is shown in Figure 31.

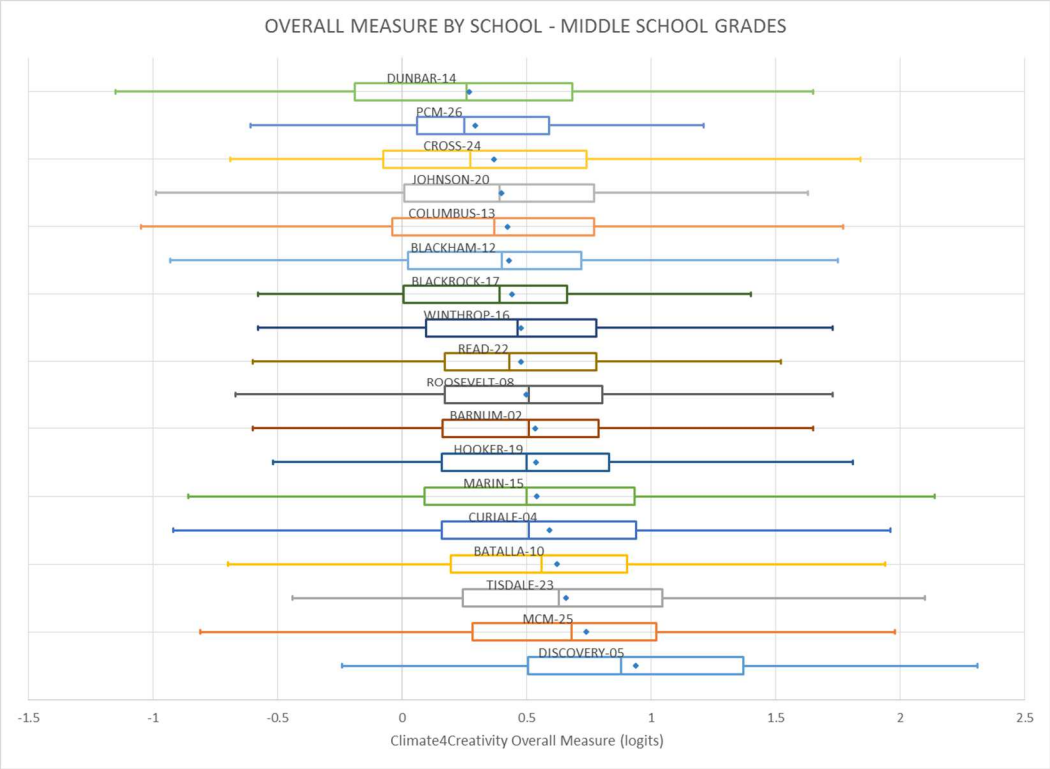


Figure 31. Comparative Boxplot of Overall Measure by School - Middle Schools.

Note. Produced using XLSTAT statistical software.

In addition to comparing schools using comparative boxplots, this analysis examined the effect sizes of the relationships between the constructs within each instrument, so as to establish the strength of the relationships between each measure at the school level. This analysis provides direct input to the theory of Climate4Creativity, which predicts the relative strength of those relationships at both the individual student and school level. Effect sizes were large in all cases, particularly in the relationships between safety and community and community and learning environments. Results of this analysis are shown in Table 28.

Table 28.
Construct Cross-correlation Analysis Findings, School Level.

		Effect Sizes: Pearson's r (r-squared, r^2)		
Instrument	**Construct**	Community	Learning	Creativity
Middle School	Safety	0.88 (0.77)	0.76 (0.58)	0.59 (0.35)
	Community	-	0.83 (0.69)	0.66 (0.44)
	Learning	-	-	0.87 (0.76)
Elementary	Safety	0.84 (0.71)	0.71 (0.50)	
	Community	-	0.90 (0.81)	

Note. [1]Combined Learning & Creativity Construct used in Elementary Instrument.

External Reference Analysis. An important source of validation evidence is evidence related to relationships with external criteria (American Educational Research Association et al., 2014). Publically available data related to the school discipline provides a credible criterion reference for use in this analysis. The Connecticut Education Data and Research (CEDaR) database (Connecticut State Department of Education, 2014), containing publically available discipline and safety data for all public schools in Connecticut, was accessed and all discipline and safety data for Bridgeport, CT schools for the 2012/13 school year were extracted to an Excel (Microsoft, 2013a) spreadsheet. All public schools in Connecticut are required to report their discipline data to the state department of education at the end of each school year, based on their mandatory record keeping, this is not statistical data, but a census of school reports based on paper and electronic records maintained by designated officers within each school. Initial analysis identified two CEDaR metrics which provided a criterion reference for overall safety and community in a school. Table 29 shows school level CEDaR data used in this analysis.

Table 29.
Actual Violence & Aggression Reports by School.

		Reports of Violence[a]		Reports of Aggression[b]	
School	**Total Enrollment**	**Count**	**Violence Ratio[c]**	**Count**	**Aggression Ratio[c]**
Barnum	625	86	0.1376	53	0.0848
CSA	310	17	0.0548	9	0.0290
Curiale	775	151	0.1948	70	0.0903
Discovery	457	9	0.0197	0	0.0000
Hall	308	11	0.0357	0	0.0000
Hallen	336	18	0.0536	12	0.0357
Roosevelt	524	136	0.2595	130	0.2481
Bryant	413	7	0.0169	0	0.0000
Batalla	1,116	100	0.0896	168	0.1505
Beardsley	377	35	0.0928	31	0.0822
Blackham	1,130	27	0.0239	7	0.0062
Columbus	877	177	0.2018	169	0.1927
Dunbar	313	278	0.8882	241	0.7700
Marin	871	231	0.2652	242	0.2778
Winthrop	686	45	0.0656	7	0.0102
Black Rock	348	0	0.0000	10	0.0287
Edison	244	30	0.1230	16	0.0656
Hooker	381	51	0.1339	34	0.0892
Johnson	819	41	0.0501	43	0.0525
Madison	506	7	0.0138	0	0.0000
Read	857	277	0.3232	138	0.1610
Tisdale	630	128	0.2032	83	0.1317
Cross	420	146	0.3476	222	0.5286
Multicultural Magnet	456	0	0.0000	0	0.0000
Park City Magnet[d]	-	-	-	-	-
TOTAL	**13,779**	**2,008**	**0.1457**	**1,685**	**0.1223**

Notes. [a]Violence = CEDaR data on *fighting & battery + acts of violence + sexual aggression + use of weapons*. [b]Aggression = CEDaR data on *threatening & verbal attacks + confrontational behavior*. [c]Ratio of reports to total enrollment at a school. [d]Park City Magnet did not report data to CEDaR for 2012/13 school year.

Counts of acts of violence were extracted from CEDaR by adding together the data for *fighting and battery*, *acts of violence*, *sexual assault*, and the *use of weapons in school*. Counts of acts of aggression were extracted from CEDaR by adding together reports of *threatening*, *verbal attacks*, and *confrontational behavior*. Counts were divided into school enrollment counts

to create a per student ratio of violence and aggression in a school. In theory, schools with higher Climate4Creativity safety scores should demonstrate lower levels of actual violence. Schools with higher community scores should demonstrate lower levels of actual aggression. Learning and creativity scores should be less correlated with actual violence and aggression. To evaluate the measures against this criterion reference, a correlation analysis was conducted between the measures for each school and the ratios computed based on CEDaR data. Correlations between aggression ratios and violence ratios and Climate4Creativity measures, by school, are presented in Table 30.

Table 30.
Relationships between Measures and External School Discipline Data.

Instrument	Measure	**Pearson's r Correlation to Violence Ratio** *(r_{viol})*	**Pearson's r Correlation to Aggression Ratio** *(r_{aggr})*
Middle School	Safety	-0.63	-0.60
	Community	-0.47	-0.52
	Learning	-0.32	-0.29
	Creativity	-0.17	-0.07
Elementary	Safety	-0.68	-0.62
	Community	-0.59	-0.52
	Learning & Creativity	-0.56	-0.45

As anticipated, the violence ratio by school showed a moderate negative correlation with the Climate4Creativity safety measure for both elementary (r_{viol} = -0.68) and middle school (r_{viol} = -0.63) instruments. The aggression ratio also showed moderate negative correlation with the safety measure for elementary (r_{aggr} = -0.62) and middle school (r_{aggr} = -0.60) instruments. These findings provide evidence that scores from the Climate4Creativity safety construct provide a valid measure of actual safety at a school: Schools with higher levels of violence and

aggression, tend to report lower scores on the safety construct, and schools with lower levels of violence and aggression tend to report higher scores on the safety construct.

Similarly, the violence ratio and aggression ratio showed negative correlations with the scores from the community measure for both middle school (r_{viol} = -0.47, r_{aggr} = -0.52) and elementary school (r_{viol} = -0.59, r_{aggr} = -0.52) instruments. These correlations were not as strong as the correlations with the scores from the safety measure, supporting the idea that these connections are not as direct. Community is related to violence and aggression, but not as closely related as is safety.

Weak negative correlations were found between violence and aggression ratios and scores for the middle school learning environment measure (r_{viol} = -0.32, r_{aggr} = -0.27), and no meaningful correlations were found between violence and aggression ratios and scores for the middle school creativity measure (r_{viol} = -0.17, r_{aggr} = -0.07). The implication of these findings is that these measures do not correlate strongly with violence and aggression, supporting the assertion that these measures really do measure something different than do the safety and community measures.

The combined elementary school learning and creativity measure, showed moderate correlations with both the violence ratio (r_{viol} = -0.56) and the aggression ratio (r_{aggr} = -0.45). This was an unexpected finding, implying that actual acts of violence and aggression may penetrate into the elementary classroom more effectively than into the middle school classroom, or stated another way, as students mature, they may become more capable of ignoring acts of violence and aggression, enabling them to more effectively isolate themselves from violence and aggression, moderating their feelings about the classroom learning environment. This finding warrants further investigation through later studies.

The second external reference data used were the 2010/11 school year results from the Comer School Climate Study (Yale School of Medicine, 2009). The Comer data were only available at a summary level, were two years out of date, and referenced a different cohort of students than did the Climate4Creativity scores. These were, however, the only available external source of data for comparison with the new measures of learning environment, since the study had been conducted across the school system two years before and reported some similar measures of school climate. The closest proxy for learning environment seemed to be the order and discipline (ORD/DIS) trait in the Comer findings. Summary findings were extracted from the Comer findings reports (PowerPoint and printed documents) and collected in an Excel spreadsheet (Microsoft, 2013a), by school, and correlation analyses were conducted between available Comer scores and scores from the Climate4Creativity measures. Correlation results are presented in Table 31.

Table 31.
Relationships between Measures and External Comer School Climate Measure.

Instrument	Measure	Pearson's r Correlation to Comer ORD/DIS[1]	Pearson's r Correlation to Comer Overall[1]
Middle School	Overall	0.06	0.09
	Safety	0.37	0.37
	Community	-0.16	-0.10
	Learning	-0.29	-0.25
	Creativity	-0.26	-0.21
Elementary	Overall	0.32	0.39
	Safety	0.51	0.54
	Community	0.15	0.25
	Learning & Creativity	0.07	0.19

Note. [1]Comer data are 2010/2011 school year, Climate4Creativity data are 2012/13 school year. Comer Middle School n = 14, Comer Elementary n = 18.

The Comer data were found to be more problematic than expected: No sample sizes, data distributions, or significance values were presented in the findings, and it was not possible to gain access to the underlying data. In addition, several schools were left out of the published findings, making it challenging to connect the new measure to the two year old Comer data. Only 14 middle schools and 18 elementary were available, and the data showed little relationship with any of the scores. The Comer metrics (ORD/DIS and School Climate) showed moderate-weak positive correlations with elementary school overall ($r_{ord} = 0.32$, $r_{sc} = 0.39$) and safety ($r_{ord} = 0.51$, $r_{sc} = 0.54$) measures, and weak positive correlations with the middle school safety measure ($r_{ord} = 0.37$, $r_{sc} = 0.37$). Community, learning, and creativity measures on the middle school instrument showed slight negative correlations to both Comer metrics, and community and learning & creativity measures on the elementary school instrument showed slight positive correlations to both Comer metrics. No significant correlations were found between the Comer metrics and the middle school overall measure.

The absence of technical quality information related to the Comer data used in the analysis make it challenging to infer meaning. For example, a general reliability score of 0.75 was reported for the DIS/ORD metric (Yale School of Medicine, 2009), but no information could be found related to reliability of the overall school climate metric, and no validation study evidence could be found for the intended uses and interpretations of the scores from the instrument.

Discussion and Conclusions. The estimates provided by the adjusted instruments used for this analysis were found to have high reliability scores, and exhibited good model fit. This study provided strong evidence for the validity and reliability of estimates given their intended interpretations and uses. The implication of this was that the construct-measure relationships,

described in the construct maps provided earlier in this dissertation, were a good reflection of the theory of Climate4Creativity. The Climate4Creativity measure has been shown by this calibration study to be a more reliable measure of school climate than its reference instrument, the Comer School Climate Instrument (Yale School of Medicine, 2009).

The core structure of the Climate4Creativity, as defined by the construct maps and items design for the Climate4Creativity Student Perspectives instruments seems sound, with one significant question outstanding: Should the general learning environment and the environment for teaching and learning creativity be treated as separate components of the model, or are they more appropriately viewed as a continuum of learning environments, with the environment for teaching and learning creativity as a reflection of a higher level of learning environment? While not the central focus of this work at the start, this issue has become a core issue for discussion in this dissertation, in the context of both the theory of Climate4Creativity and the theory of action advocated by this work.

At an individual student and schoolwide level, this study supports the theory of Climate4Creativity, in that the hypothesized relationships between the constructs were demonstrated empirically for both elementary and middle school students. Note that the relationship structure for elementary students was simplified by the merger of the learning and creativity construct into a single unified construct of the learning environment. Figures 32 and 33 show the empirical strengths of the relationships between constructs for middle school and elementary school students. This supports the core idea that safer schools with stronger, more caring communities provide individual students with better learning environments, and that general learning and learning creativity are intrinsically linked in the minds of students in public elementary schools, even though these students may not always name these components as such.

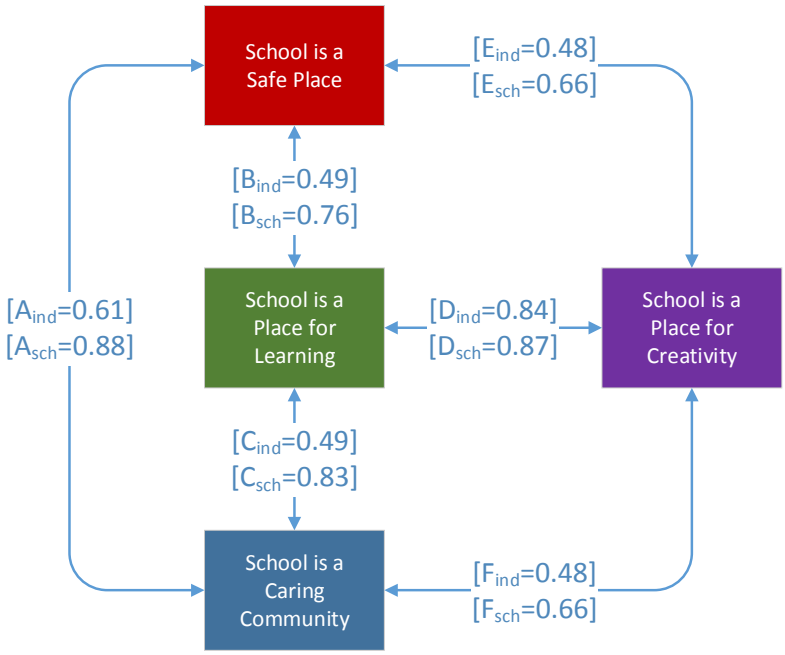


Figure 32. Empirical Construct Inter-relationships, Middle School

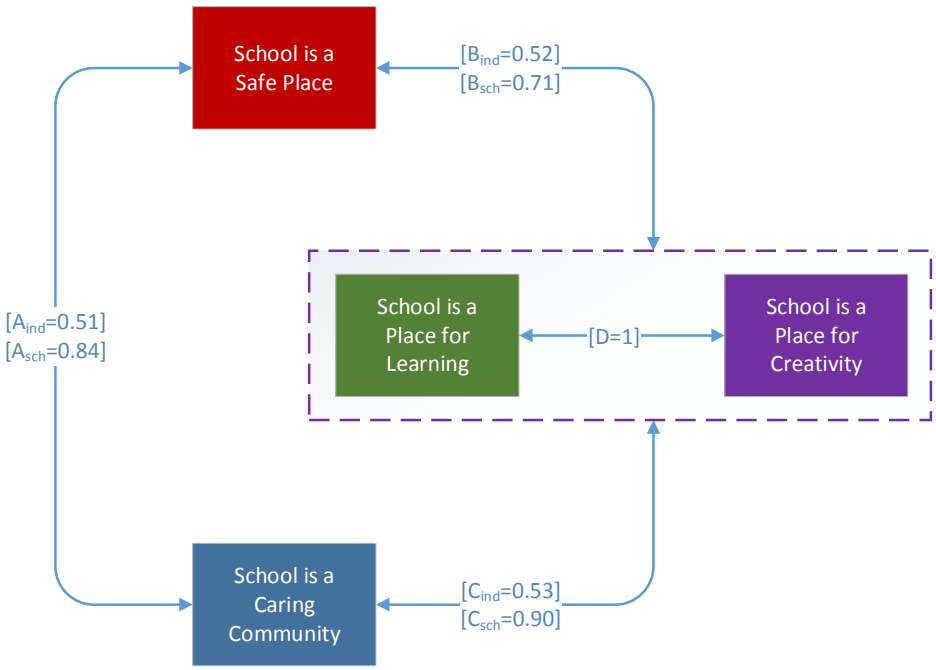


Figure 33. Empirical Construct Inter-relationships, Elementary School

The initial structure of the theory of Climate4Creativity included four major constructs. The first two constructs, *School is a Safe Place* and *School is a Caring Community*, based on this study, are sound and supported by the empirical data from this significant study. The third and fourth constructs, *School is a Place for Learning* and *School is a Place for Creativity*, are sound and supported at the middle school level, but were merged into a single construct, *School is a Place for Higher Order Learning*, due to the performance of the measure at the elementary school level. This raises an important structural question about these ideas: Should the general environment for learning and the environment for creativity be separate or should they be merged into a continuum, with creativity being the highest level of learning, within a single construct concerned with the learning environment? At the elementary level, it is clear that they should be merged into a single learning environment continuum. The question of whether the same would be true for a merged middle school construct warranted an additional analysis using the empirical data from this study.

Noting, first, that the learning and creativity constructs in the middle school instrument were highly correlated ($r = 0.87$), it was clear, even from the scatterplot shown in Figure 29 that these constructs performed in a similar way. Higher level learning environment schools *tended* to have higher level creative environments, too. Qualitatively, it can be argued that if New Bloom's Taxonomy (Anderson & Krathwohl, 2001) is a good model for learning objectives, the constructs should be merged to better reflect the levels of learning found in the taxonomy. In order to assess this, a psychometric analysis was conducted using a merged learning and creativity construct for middle school age students. The newly merged construct performed better psychometrically than the two separate constructs. Reliability was significantly higher for the new merged construct ($r = 0.92$, $\alpha = 0.97$), and only one misfitting item was identified (23-

FGRD: *what grade to you expect to earn in your favorite class?*). The misfit in this case is likely due to the very low number of responses at category 4: *D* or category 5: *F* (less than 3%). This is qualitatively not surprising, since few students would expect to fail their favorite class. When this item was removed from the measure, reliability was unchanged and all items fit the model.

No item polarity issues were identified and item distribution was as expected. As expected, it was more difficult for students to select high scores on *least favorite* subject questions than on *favorite* subject questions, implying a qualitative relationship between the learning environment and favorite subject responses. Essentially, students feel that the learning environment is better in their favorite subject than in their least favorite subject. This supports the theory of Climate4Creativity as expected by the model. The merger of these constructs led to zero additional rows of missing data in the cleaning process, but three additional extreme low scoring rows and five additional extreme high scoring rows were removed, leading to a new clean record count of 3,111.

Individual level cross-correlation analysis showed similar properties to before, with the safety and community cross-correlation effect sizes (Pearson's r values) computing as 0.46 and 0.47 respectively, and the merged learning and creativity construct showing a high level of correlation (Pearson's $r = 0.81$) and coefficient of determination (r-squared, $r^2 = 0.66$) when compared with the overall measure at the middle school level. School level cross-correlation analysis also showed similar properties, with Pearson's r values for correlation effect sizes with safety and community of 0.71 and 0.81 respectively. The merged learning and creativity construct showed a high effect size ($r = 0.89$; $r^2 = 0.79$) when compared with the overall measure. The boxplot analysis showed that the new merged learning and creativity construct performed almost as well as the two constructs individually, although the merged construct did

not differentiate between schools quite as well as the individual constructs. The new merged construct performed similarly to the two constructs individually, in comparison to external reference data. No major psychometric or qualitative issues emerged related to the merged construct, it performs as well as the individual constructs, with a higher reliability than the individual constructs. Based on the use of the new merged construct, the middle school student perspectives on the theory of Climate4Creativity would be as shown in Figure 34.

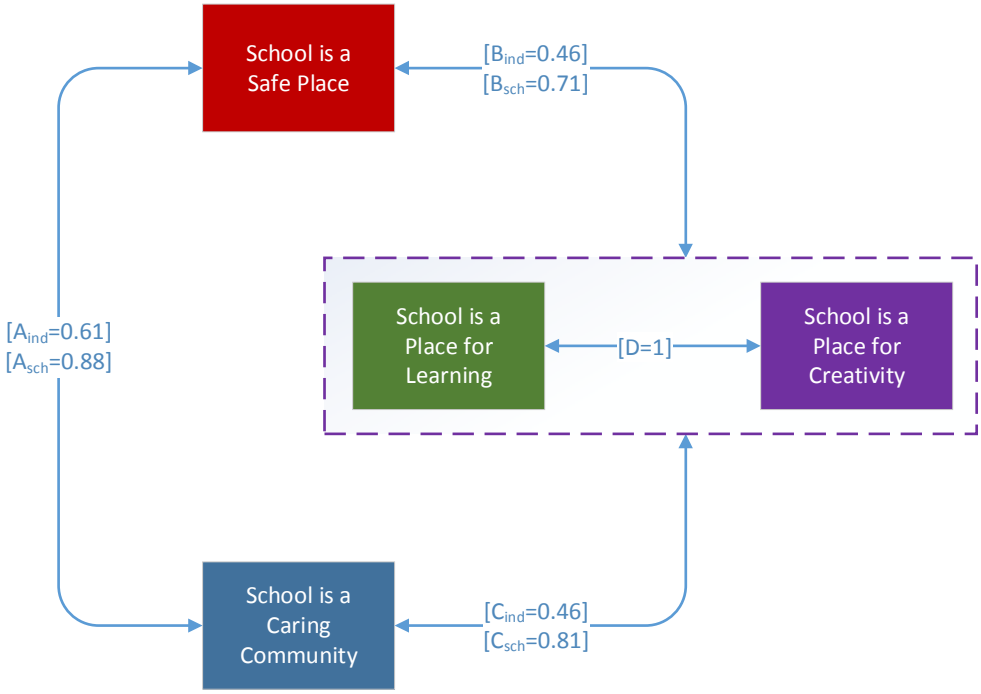


Figure 34. Revised Empirical Construct Inter-relationships, Middle School

This adjustment to the construct model better reflects the empirical data, without significant loss of detail or meaning in the Climate4Creativity Student Perspectives instruments.

While an equating study was not completed as part of this dissertation, the majority of items in the elementary and middle school measures were common. Within the safety items design, only 2 out of the 28 items were not common, and within the community items design,

only 1 out of the 18 items were not common. The combined learning & creativity items design differed between the elementary and middle school instruments in a more significant way: All middle school items were asked twice – once for the respondent's favorite subject and once for the respondent's least favorite subject; elementary school items were asked once, since elementary school students typically do not have different classroom environments for each subject. The 18 elementary school items were asked twice in the middle school items design, and there was one additional middle school item (asked twice). It seems likely that the single learning & creativity measure yielded by the elementary school instrument would fall between the least favorite and most favorite subject learning & creativity scores yielded by the middle school instrument. Figure 35 illustrates the shifts in perception between the Climate4Creativity measures.

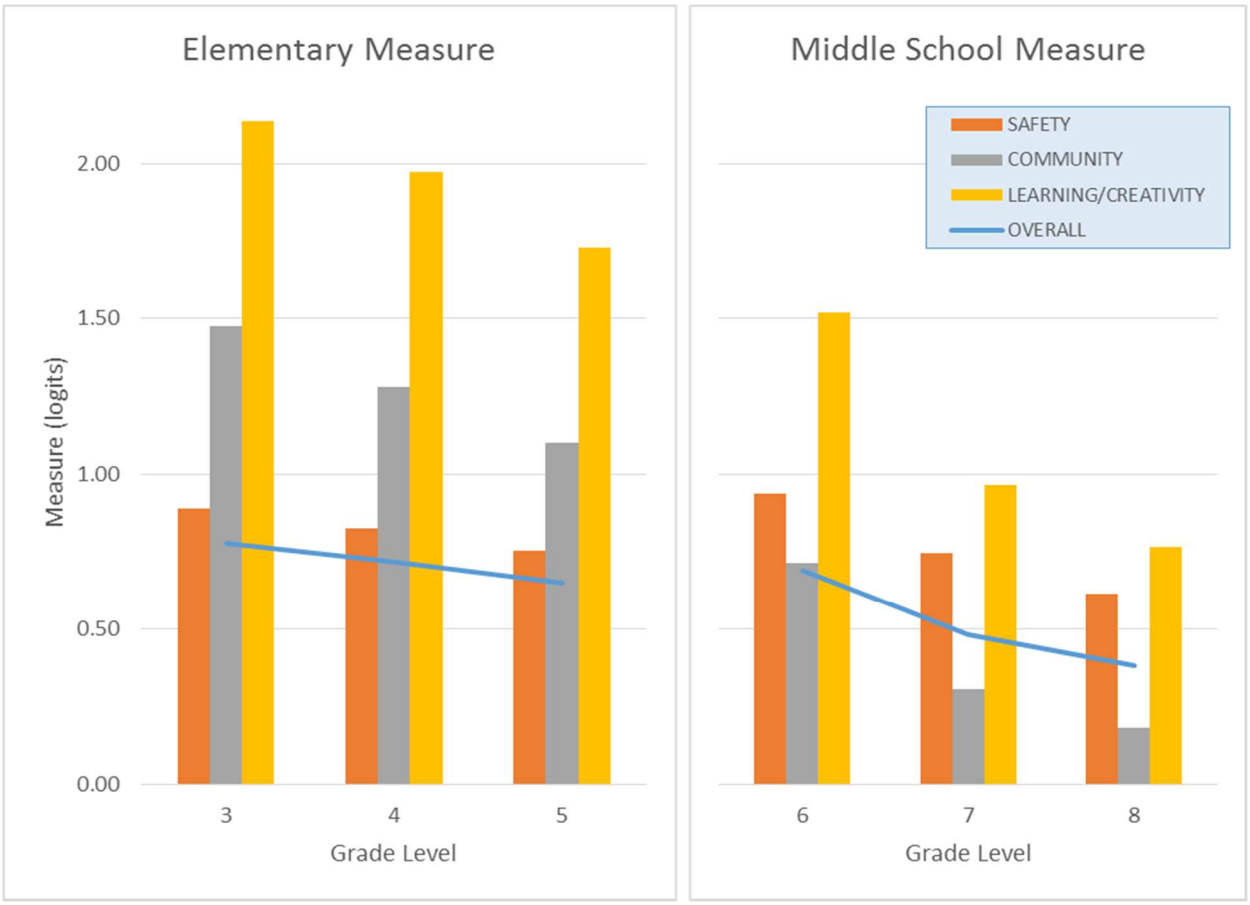


Figure 35. Climate4Creativity Measures by Grade Level.

Student perceptions of school climate overall and of physical and emotional safety appear to remain fairly constant through the grades. This is likely because as young people develop their sense of independence, they begin to *toughen up* to activities which in the earlier grades might be considered unsafe. Risk taking behavior does increase with age through the middle school years. Perceptions of community tend to deteriorate through the grades, once again, perhaps a reflection of both emerging independence, and of deliberate school policies designed to push students to develop self-reliance as they grow up. The learning and creativity measure tends to deteriorate significantly from the elementary grades to middle school grades. This implies raised expectations and increased critical judgement of self and others (particularly with

respect to creativity), an increase in variability in the data due to more variety in classroom settings and teacher practices. Third graders typically have a main "homeroom" classroom and teacher, whom they get to know very well throughout the school year, whereas eighth graders typically travel to different classrooms throughout their school days and interact with a variety of teachers, who may create very different learning environments.

One important question is the impact on the measures of personal experience of bullying and aggression in school. By extracting the control variable, *have you personally experienced bullying at your school?* and comparing average measures based on students who selected each option, we can readily see that students who report *being bullied now* report the lowest levels of safety and community, followed by those who *have been bullied in the past*. Students who have *not personally experienced bullying* express high levels of safety and community, and students who report a belief that *no bullying exists in their school* report the highest levels of safety and community (Figure 36).

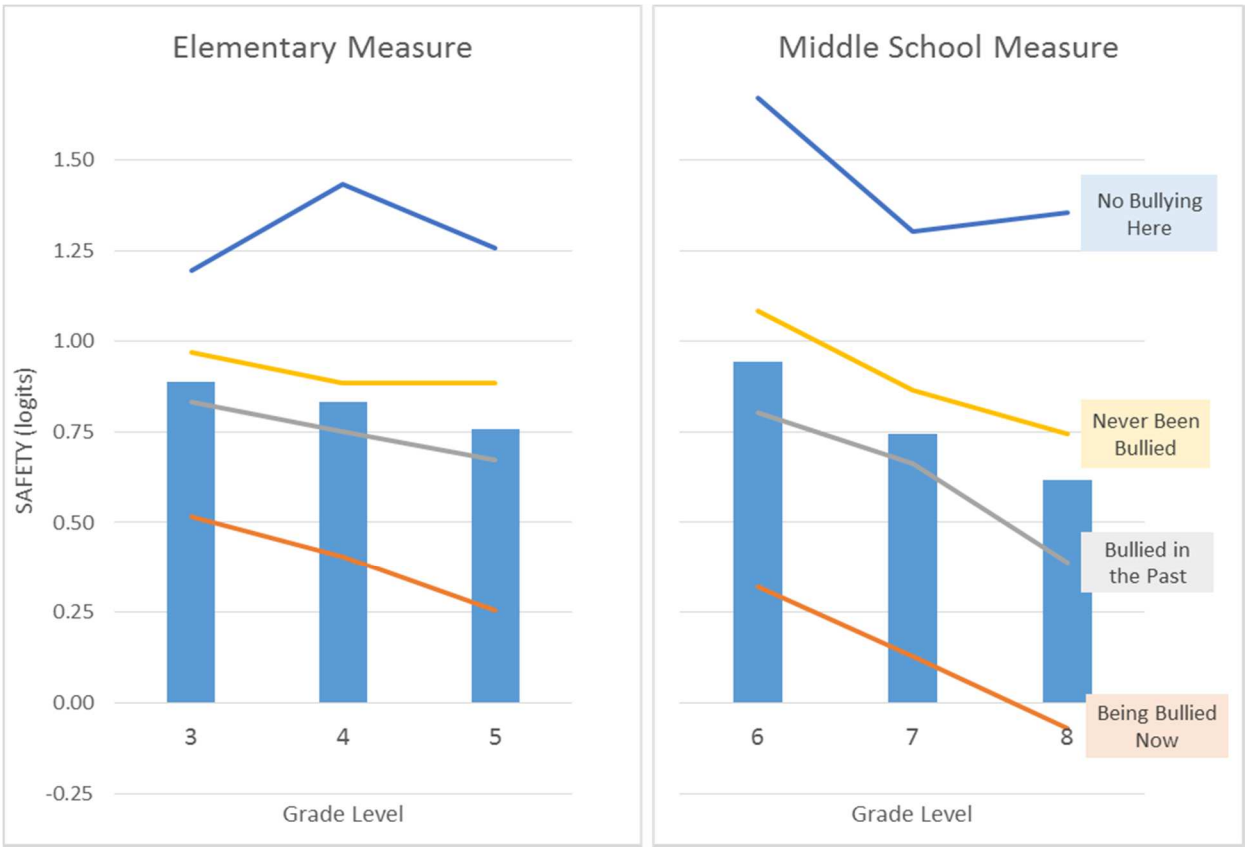


Figure 36. Safety Measure by Personal Experience of Bullying.

This is not as clear when we examine the learning and creativity measure beyond the 6th grade (Figure 37). In 7th and 8th grade, personal experience of bullying seems to have much less impact on the learning environment than in earlier grades. This may be due to emerging toughness, or to the pervasive nature of bullying, whereby, almost all 7th and 8th grade students have experienced bullying at some point in their school career, it may also be due to the inability of lower grade students who are victims of bullying to *escape* their classroom, since these students typically remain in the same classroom with the same teacher for much of their school day.

The implication of this is that bullying experience may be a powerful, pervasive determinant of school climate and feelings of safety and community in all grades, but does not

penetrate as strongly into perceptions of the classroom learning environment to the same degree, particularly at the higher grades. This finding triangulates with the external reference analysis using CEDaR (Connecticut State Department of Education, 2014) data. Table 30 shows that the prevalence of violence and aggression penetrates into the elementary school learning environment at significantly higher levels (r_{viol} = -0.56, r_{aggr} = -0.45) than it does into the learning environment (r_{viol} = -0.32, r_{aggr} = -0.29) and creative environment (r_{viol} = -0.17, r_{aggr} = -0.07) in middle school.

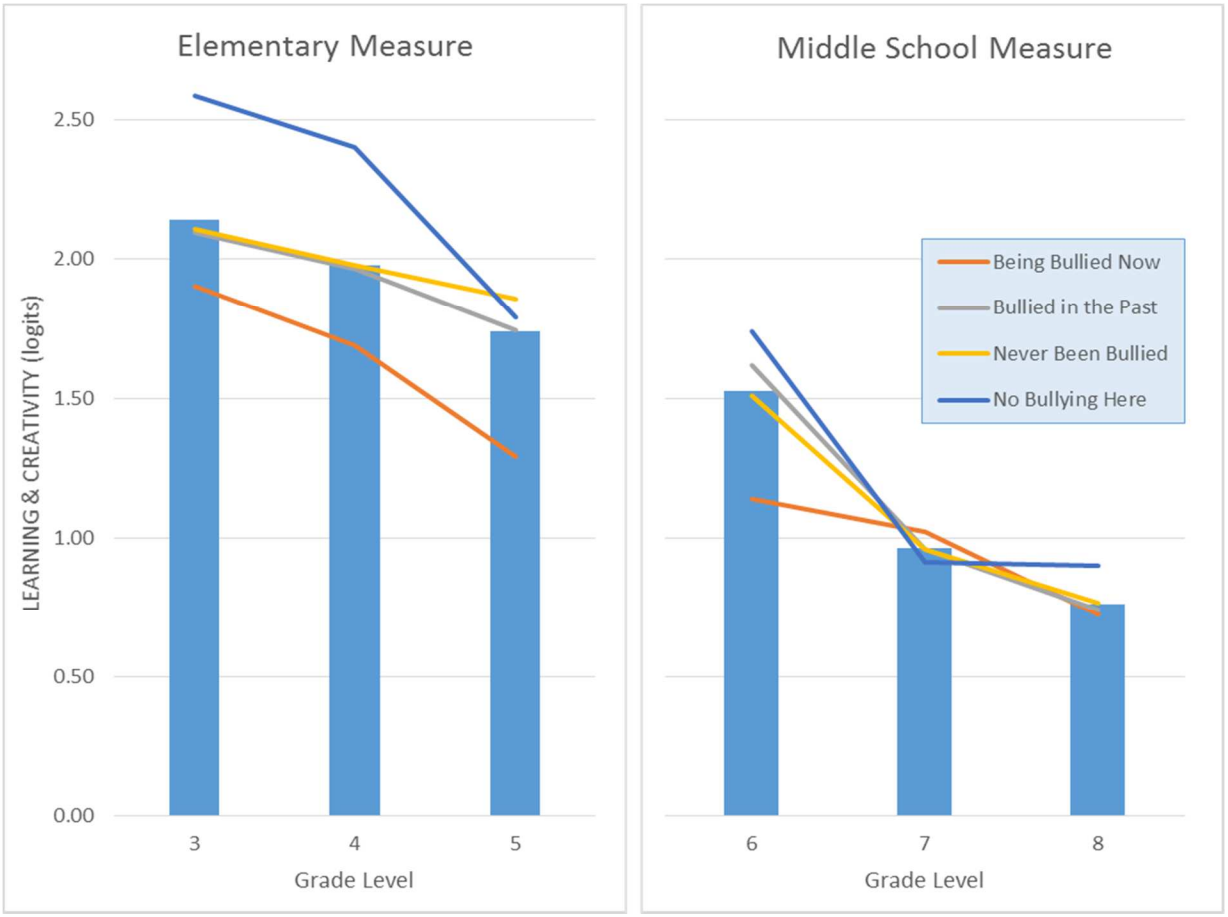


Figure 37. Learning & Creativity Measure by Personal Experience of Bullying.

Based on these findings, the Climate4Creativity construct model will be changed to reflect a new combined, renamed construct called *School is a Place for Higher Order Learning.*

This new construct name reflects a combined, continuous measure of the learning environment, reflecting New Bloom's Taxonomy (Anderson & Krathwohl, 2001), and uses a combined items design based on the two original items designs.

Limitations and Next Steps. The first limitation of this study was that it included only schools in the Bridgeport, CT school system, one of the worst performing school systems in the U.S., in terms of economic and ethnic achievement gaps (Hemphill & Vanneman, 2011), and a school system with a reputation for serious issues related to safety, school community, and the learning environment. To test the higher levels of the measure, a test of the Climate4Creativity estimates in prestigious private schools or creative learning schools, such as a Montessori or Steinar schools, would be warranted. In line with this, some additional items might be added at the higher end of the measures, to allow differentiation between students with higher levels of the measure.

Second, in a subsequent study, the elementary and middle school measures should be linked through a formal equating study. This will allow for a more robust analysis of trends and shifts in student perspectives based on equating the two measures.

Third, the instruments should be translated and tested in Spanish, and perhaps other languages, so as to engage dual language learners. In addition, an early childhood and high school age version of the instrument should be constructed and tested. These versions should be constructed to provide a continuum of measures tracking classroom and school Climate4Creativity from PK-12. An undergraduate measure could also be developed, enabling long term cohort studies using a PK-16 developmentally appropriate measure.

Fourth, additional external reference checks should be completed along with the next major study using these instruments. Specifically, an established observational learning

environment measure and an established observational antibullying measure should be used in all classrooms within one school, ideally at the same time as an integral part of the major study. These observational results should then be compared to the Climate4Creativity measures using a multi-trait multi-method matrix.

Finally, several detailed items in the data from this study warrant additional analysis, and may lead to interesting and useful insights for school leaders and educators. Examples include: 1) An analysis of the core differences between responses for the favorite classroom and least favorite classroom could yield interesting insights into how to develop better, more engaging classrooms; 2) An analysis of why those who have personally experienced bullying at school might lead to interesting ideas about how to better prevent and minimize the impact of bullying behaviors in a school; 3) An analysis of the middle school questions about seeing drama, seeing bullying, and experiencing bullying might lead to insights about how to better identify and address negative social interactions among teens and preteens in school contexts, with particular emphasis on identifying precursor behaviors or precursor environmental factors which could be used as an early warning system for bullying behavior; and, 4) An analysis of student perspectives on how difficult or easy their school work is, combined with the measures of learning environment and community, might lead to some interesting insights in the area of curriculum design and lesson planning. This is a rich and diverse dataset providing the opportunity for extensive additional data analysis and potential development of strong insights in the fields of school leadership, antibullying practice, school community building, curriculum and lesson planning, and teaching practices.

CHAPTER EIGHT:

DISCUSSION & CONCLUSIONS

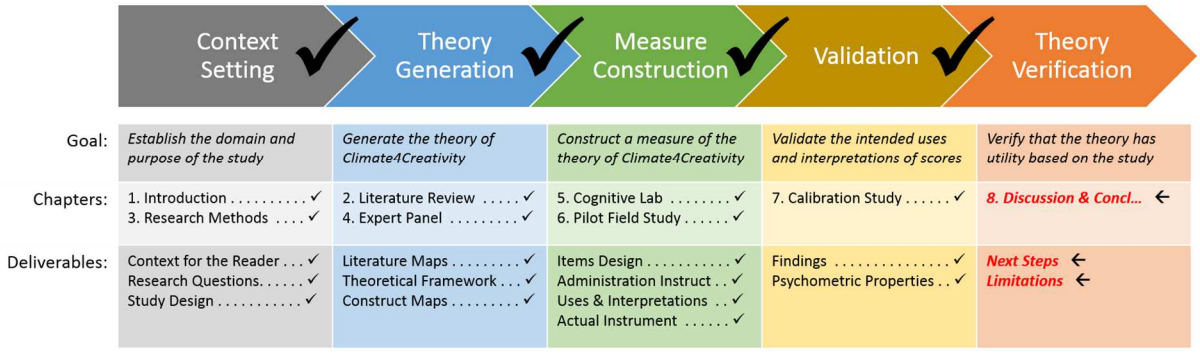


Figure 38. Dissertation Roadmap – Discussion & Conclusions.

By enhancing the practice of school improvement with a specific focus on *teaching and learning creativity*, an opportunity emerges to better prepare young people to participate in the social life of the planet, thereby empowering the sustainability of human society. To achieve this, an action-oriented *theory of climate for creativity* was proposed in chapter 4, a measure of the theory was constructed in chapters 5 and 6, and the scores from the measure were calibrated in chapter 7. An argument for the validity of the intended uses and interpretations of the scores from the measure is developed in this chapter.

Validity

Validation of Intended Interpretations and Uses. Validation refers to the intended uses and interpretations of the scores yielded by a measure, not to the measure itself. This simple, yet powerful statement is fundamental to the process of evaluating the validity of the uses and interpretations of scores from a measure such as the Climate4Creativity Student Perspectives

instruments. The intended uses and interpretations of the measures were defined explicitly at the end of chapter 6 of this document, along with the items design, following completion of the two pretest studies, study II (cognitive lab) and study III (pilot field test). The interpretive validity argument presented here is based on these intended uses and interpretations of the scores yielded from the two instruments described in this dissertation.

Validation Evidence Based on Test Content. The first stage in the development of these instruments included an expert panel study where seventeen diverse content experts from teaching, creative practice, school leadership, positive behavior interventions and supports, and educational psychology provided detailed input to the construct design and measurement structure, as well as providing input to the intended uses and interpretations of the instrument. These experts included teachers, artists, parents, school professionals, special education professionals, and professors from an NCATE accredited teacher education program. A detailed breakdown of the qualifications and experience of expert panel participants demonstrated their expertise in the domains covered by the measure.

The analysis of the shifts in construct measures by grade level and by personal experience of respondents with bullying, presented in chapter 7, supports the measure as a sound measure of safety and community, since actual experience of bullying could reasonably be expected to lead to a lower feeling of safety and community in a student, and higher prevalence of bullying experiences in a particularly school would likely imply a lower level of safety and community in that school. While grade level analysis shows developmental shifts in the higher order learning construct, the bullying experience analysis provides an unexpected response which implies that bullying experience in the past does not significantly affect perception of the environment for higher order learning in the later grades (particularly 7^{th} and 8^{th} grades). This is an interesting

finding which warrants additional investigation in later studies, particularly when controlled through a formal equating study, linking the developmental stages of the elementary and middle school instruments.

The literature directly supports the structure of the instruments, and the thematic structure of the literature as presented in chapter 2 of this dissertation provides evidence of this direct support. Several counter theoretical studies are discussed in the review of the literature, supporting the perspective that this literature review was thorough and systematic in its selection of studies for inclusion. The items included in the items design are cross-referenced to the literature in Appendix E, along with a detailed definition of the items. This is a useful presentation of the theoretical validity of the content of the measure.

Validation Evidence Based on Response Processes. A carefully constructed cognitive lab study was conducted as part of the pretesting processes for these instruments, providing detailed input from target respondents, including a study of the cognitive processes used in responding to items in the measure. Study participants were students from grades 3, 4, 7, and 8, drawn from the actual target population the measures were intended to address. This formed an excellent representative sample in terms of gender, ethnicity, and age. Findings from the study were clearly integrated into the instrument design, particularly with respect to the wording of individual items and response options.

In addition to the cognitive lab study, a relatively large pilot field test of the instruments (n = 285) provided interesting response-based input to the items design. For example, the use of the wording, *tell me to work hard at school* versus *encourage me to work hard at school* in the measure was identified as a potential issue through analysis of quantitative data from the pilot test. In this example, the word *tell* was used in the elementary instrument and *encourage* in the

middle school instrument. Poor model fit was observed for this item in the elementary school instrument, but not in the middle school instrument. The test developer concluded that some younger respondents were interpreting the word *tell* as a negative response, rather than the intended interpretation. The wording of this item was changed to *encourage* and the issue was eliminated, as demonstrated by the good model fit found in the calibration study.

Validation Evidence Based on Internal Structure. The pilot field test study assessed the individual items in the measure, and the measure overall, and made careful refinements to individual item wording, and the overall item structure, using the empirical evidence from the study to drive good decisions. Calibration study findings included comprehensive evidence of reliability of the instruments, showing that they at least matched existing reference instruments being used in the field. A number of internal structure analyses based on the calibration study demonstrated that the instruments performed well overall and were properly aligned with respondent cognitive processes. This was particularly clear with the safety and community constructs. A detailed discussion of the learning and creativity constructs presented in the calibration study concluded that these constructs should be merged into a single higher order learning construct, and sound empirical and theoretical evidence was presented for this conclusion.

The Rasch measurement model, used to analyze the instrument, provided strong evidence of validity. Every item in the finalized instrument exhibited mean square infit and outfit values in the range 0.5 through 1.7, showing that the instrument performed according to model expectations. Item polarities were examined, and found to support item-measure correlations. Analyses of the empirical category functioning, using an assessment of the Andrich thresholds for the overall measure and for the safety and community constructs found these categories to

function appropriately. As part of a comprehensive assessment of the learning and creativity constructs, category functioning was found to be an issue, and categories were collapsed using appropriate procedures. These constructs were combined into a single higher order learning construct, yielding improved psychometric properties for the newly merged construct.

Analysis of school level results in the calibration study showed that the measure was capable of discriminating between classes of schools, particularly at the higher end of the measures. One school in the calibration study, Discovery Interdistrict Magnet School, significantly outperformed almost all other Bridgeport schools on all measures. This particular school is perceived within the district as the best school in the district. This perception aligned well with empirical findings from the measure.

During the calibration study, differential item functioning tests were performed on the overall measure for both instruments, and found no non-negligible evidence of gender or ethnic bias within the empirical data. This is to be expected, since the cognitive lab study performed early in the instrument development process included a representative sample of target respondents, and item and answer choice wording was carefully refined based on the cognitive processes employed by target respondents.

Validation Evidence Based on Relations to Other Variables. Moderate negative correlations were found between the safety and community measures and publically available school discipline data extracted from the State of Connecticut's school discipline database, providing strong evidence that these components of the instrument accurately measure levels of physical and emotional safety in a school, because schools with higher measures of safety and community should tend to have lower incidents of violence and aggression per student, and vice versa. The Comer School Climate Survey (Yale School of Medicine, 2009), administered in a

subset of study schools two years prior to the calibration study might have provided supporting evidence for the learning environment construct, but in fact did not. It is likely that the timing difference and the level of recent change in Bridgeport schools invalidated this evidence. A valid and reliable learning environment measure should be implemented in parallel with the next major study, in the same school, at about the same time, so as to continue to build evidence based on relations to other variables. A multi-trait multi-method matrix (MTMM) showing the relationships between the Climate4Creativity Student Perspectives measures and other external measures should be presented at that time.

Evidence for Validity and Consequences of Testing. Outside of the scope of this research work, a team of internal experts within the Bridgeport schools created a draft Intervention Action Planning guide to help school leaders take effective action based on the use of the Climate4Creativity measures. This process, facilitated by the Positive Behavioral Interventions and Supports (PBIS) team leader for the district, while not directly part of this dissertation, provides some evidence of the connection between the measures and the consequences predicted for each measure. In addition, expert panelists engaged in the early stages of the study provided specific input on the applicability of the measures to action based on findings, and concluded that the measure would likely have high utility as a driver of focused improvement action in a school.

Additional Considerations. The implementation of this measure was well organized, and easy for schools to implement. The research team did not need to physically visit schools or be in classrooms, no classroom teachers needed to be trained, and parental consent was easy to obtain since one of the major intended uses of the measure was to provide state-mandated school climate assessment data to the state department of education. Training for administrators and

onsite coordinators was well received, with excellent feedback provided by trainees, and documentation was clear and easy to use. In the case of this study, the instrument was provided free of charge to the school system, and pretesting was partially funded by a research grant from Phi Delta Kappa, the honor society for professional educators. Going forward, it is unlikely that administration of the instruments will continue to be free to schools, but, it is expected that the system will evolve into a low cost, low hassle, highly effective option for school leaders.

Integrating the Validation Evidence. The scores yielded from high fidelity administration of Climate4Creativity Student Perspectives Instruments (C4C/SPE and C4C/SPM) provide critical, valid input into four major uses. All the measures exhibit adequate reliability for use in schoolwide, classroom, and parental decision making. Cronbach's Alpha reliability scores in excess of 0.90 for all measures, and Winsteps modeled, error adjusted, person separation reliability coefficients in excess of 0.75 for all measures.

The first intended use of the instruments is to support strategic school improvement, by providing benchmarking inside the school and by helping school leaders to focus their school improvement efforts on the most significant constraints to school improvement. In order to act as an effective benchmark, the measures must exhibit stability of use over time, in other words, they must show resilience and consistency across multiple uses. The high reliability scores computed from the calibration study in twenty five Bridgeport, CT schools provides strong evidence that the scores from the measure will reliably report the current state of safety, community, and the environment for higher order learning in a school. Beyond benchmarking, use of the measures to focus strategic improvement efforts in a particular school is supported by the analysis presented in the calibration study, demonstrating that schools can be effectively differentiated by the measures. The theory of constraints model for the management of school

improvement is supported by substantial evidence in the operations research literature and in the professional practice of project management. In this *thinking model*, the primary purpose of this type of measure is to focus school leaders on a small number of strategic actions, designed to improve specific critical success factors in a carefully planned process of ongoing improvement. There is evidence in the calibration study to support the theory of action proposed in this dissertation, however, a long term study should be implemented to include multiple uses of the measures and extensive use of theory of Climate4Creativity to drive leadership action, in order to fully evaluate the validity of this theoretical claim *in action*.

The second intended use of the measures is to support schoolwide action on safety and community and classroom level tactical action to improve the learning environment. This intended use first specifies that the responses to individual items within the instruments may be used to guide tactical actions to improve schoolwide safety and community, provided the sample size for the individual item or set of items is sufficient. Care must be exercised by the user when taking action based on individual items, but, where a group of items point toward a particular recommended action, the evidence provided by the calibration study strongly supports use of groups of items in this way.

The third intended use of the measures is to support accountability and parental choice by reporting summary safety and community measures to state agencies and Climate4Creativity measures to the public. The Connecticut department of education provided a specific questionnaire to Bridgeport schools in 2013, and several school leaders were able to use the Climate4Creativity measures to satisfy many of these reporting requirements. This actual use of the measures from the calibration study by Bridgeport school leaders provides evidence in support of this intended use of the measures for this purpose. Summary data from the calibration

study was shared with the school superintendent and school board members, and was well received.

The empirical evidence from this program of research clearly supports these three intended uses and interpretations of the scores from the Climate4Creativity estimates.

Discussion

Dystopia in Education. Writing in *Amusing Ourselves to Death* (1985), Neil Postman compared the dystopian visions of Huxley's *Brave New World* (1932) and Orwell's *Nineteen Eighty-Four* (1949). He described two very different, but equally evil strategies in which citizens were oppressed: Orwell's vision of an evil dictatorship violently removing the rights of individual citizens and using pain and suffering to keep the population under control, and Huxley's vision of a benevolent dictatorship eliminating pain and suffering through programs of hedonistic pleasure and citizenship education to achieve *exactly the same ends*. In Orwell's vision, the citizens knew they are being oppressed, but could do nothing to change it. In Huxley's vision, the citizens did not even know they are being oppressed. In *Brave New World Revisited* (1958), Huxley analyzed whether the world had, in his view, moved toward or away from this dystopian vision. He concluded that society had, through the defeat of Nazi Germany in World War II, moved away from an Orwellian future, but had simultaneously moved toward his own dystopian vision of a population kept distracted and stupid by a culture of increasing edutainment and simultaneously oppressed by their own consumerism.

Today's industrial model public schools are not, of course, Huxley's dystopian vision made real. Instead of the alpha, gamma, and delta populations born, bred, and educated in Huxley's *London Hatchery and Conditioning Centre*, we have educational achievement gaps (Hemphill & Vanneman, 2011) created accidentally by economic, demographic, geographic, and

societal pressures. Our own dystopia was not deliberately created by a benevolent dictator, but was created by well-meaning policy choices made over the decades. Standardized testing was intended to raise standards for everyone. Common core was intended to help students be better prepared for their future careers. The professionalization and standardization of teaching practices was intended to make teachers more effective at delivering standardized content across the country. Positive behavior interventions and supports (PBIS), social-emotional learning (SEL) systems, educational psychology, medication of young people, were all intended to make classrooms more effective and efficient, so that students could continue to learn. The commercialization and politicization of texts was intended to make the texts better, more accessible and more usable by students. All of these strategies were intended to promote a sense of being part of a nurturing community, feelings of belonging and group identity, and long term stability for society: "COMMUNITY, IDENTITY, STABILITY" (Huxley, 1932, p. 2).

Ironically, in a postmodern world where information is more accessible and ubiquitous than ever before, it is more difficult than ever to make sense of the information. There is simply too much information, too many choices, and too many conflicting sources. According to Kurzweil (2005), the rate of paradigm shifting innovation approximately doubles every decade, and the amount of recorded human knowledge approximately doubles every year. This unprecedented explosion of new ideas and new knowledge has created an environment of rapid change and exponential growth in complexity. This global reality, combined with decades of well-meaning but misdirected efforts to reform our public education system, combined with the planetary reality of global warming, of emerging resource limits, of radicalization and war, has created a unique and terrifying situation, a unique and potentially dystopian future of our own.

The Climate4Creativity Antidote. What, then, is the antidote to this dystopian future? In my view, an important part of this antidote is educating public school students and future citizens for creativity and creative thinking. Society cannot fix the future with more of the same thinking that created the present. Working harder at what doesn't work will not magically make it start working. Caught in the tyranny of the squeaky wheel, society must shift laterally and apply lubrication in the right places and in the right ways. The Climate4Creativity framework begins with school as a safe place, builds a strong, caring community wrapped around and inside the school, and then brings classrooms filled with positive learning environments, environments focused on teaching and learning higher order thinking. These positive classrooms are filled with excitement, fun, and engaged learning. All students are engaged, whether creative or non-creative learners. Non-conformity, emotional instability, loneliness, singular focus are embraced. Creative individuals are able to transcend *normal* learning and find the special, *deep* meaning in everyday things. Creativity is the inverse of standardization. Without creativity, we would inhabit a very dull planet. Without creativity, humankind would never have evolved. The real question is not, in fact, the value of creativity and creative thinking. That is a given. The real question for today is, how can school leaders actually bring creativity and creative thinking into our public schools? The Climate4Creativity framework is a foundational part of that solution.

The Five Boats Analogy. In the lexicon of theory of constraints, there is an anecdote called *the five boats analogy* (Leach, 2004). In this myth, five ships captains arrive in harbor on the same day and demand that their cargos be unloaded by the five longshoremen waiting on the dock. The harbormaster has a difficult choice: Give in to the five captains and start unloading everyone at once, or, choose one boat to unload first? A desire for fairness, fear of making one

captain feel slighted, and intuition about getting work started on time, all conspire to tell the harbormaster to send one longshoreman to each boat and make all five captains happy by beginning the unloading process for all five boats at once. In this solution, five days later, all five boats are unloaded, and everyone seems happy. But, what if the harbormaster had, in fact, decided to select one of the boats at a time, and send all five of his longshoremen to one boat at a time, unloading each in turn. The first boat would have been completely unloaded that first day, the second would have been completely unloaded on the second day, and so forth. The fifth and final boat would have been completely unloaded on the fifth day. In this scenario, the fifth boat would have been no worse off that in the, *everyone starts together* scenario, but every other boat would have been unloaded sooner, and therefore could set off on new adventures earlier, freeing dock space for additional boats. The average unload time would have been cut in half. The only conceivable downside of this approach is that some of the captains, particularly the fifth captain, might have felt slighted and become irrationally upset, even though he would have been in no worse position than before.

This analogy powerfully illustrates the idea that systems have goals, and that focus on these goals drives actual performance. In the case of a harbor, the goal of the system is to unload all ships as quickly as possible. In the five ships analogy, the best way to do this is, in fact, to pick one project and see it through to the end first, rather than try to achieve all projects simultaneously. Singular focus on what matters most to the system goal is at the heart of theory of constraints, and is at the heart of the call to action in this dissertation. In today's education system, the singular transcending paradigm (Meadows, 2008) for all teaching and learning is improving standardized test scores. Everything else in the system aligns itself around that transcending paradigm. The transcending paradigm of public education needs to change. Then

the goals, organization, rules, information flows, feedback loops, and other system components will realign themselves to balance standardized test scores. This shift from focus on standardized test scores to a more balanced paradigm is, in my opinion, the five boats analogy for the future of public education.

Where This Study Fits. This program of research has gone deep into just one aspect of what it will take to transform public schools into places where young people are prepared for work and higher education, and simultaneously prepared to become active participants in U.S. democracy. Of course, schools in every community in the U.S. find themselves in different circumstances, with different leadership, different histories, different resources at their disposal, and with different challenges and barriers to be overcome. This work calls for and supports a strategic transformation of U.S. public schools to become places purposefully constructed to encourage and enable creativity and creative thinking. This transformation will, of course, be challenging on many levels.

The immediate value of this work is twofold. First, by creating a credible definition of the Climate4Creativity, educators and school leaders have the means to clearly articulate a cohesive strategic vision for their school. Second, by implementation of a credible measure of the Climate4Creativity, school leaders have the means to measure the current state of their school, enabling them to take focused action, and measure the value of their action over time. These two capabilities provide the benchmarks and guideposts for a process of ongoing improvement in any public school.

Development of a Theory of Action. A theory, whether verified or not, provides a way of *understanding* a situation, based on experience and evidence. A theory *of action* moves beyond understanding a situation, and provides a way of *acting* in a situation, understood

through the underlying theory but designed to *change* the situation in some meaningful way. Use of such a theory of action is a pragmatic approach to implementing strategic change in complex institutions, such as schools. Strategic goals are expressed as *outcomes* from an overall strategy, and will often seem overwhelming to individual leaders and professionals, particularly when the goals are transformational in nature. It is hard for humans to see the path from where they are today to where they need to be in the future. Clearly, in the case of the goals of this study, success would involve moving beyond merely understanding school climate and the environment for teaching and learning creativity in a school to encompass transformation of many mainstream U.S. public schools. It is important to connect the transformational goal of bringing teaching and learning of creativity into an elementary school, with the concrete steps needed to actually implement the goal in actual practice. Organizations cannot implement transformation in one big step, organizations must evolve over time, employing small but focused tactical moves, in the context of the wider strategic agenda (Duck, 1993). This allows change to be implemented effectively, while limiting the negative impacts of too much change. Public schools, in particular, need to continue to deliver learning to their current students, while implementing structural changes to improve themselves over the long term.

A useful theory of action, then, includes a deep understanding of the environment being changed; an honest, realistic analysis of the capabilities and resources at the disposal of the organization; a clear logic of the change; and, valid measures of success. Organizations break their strategic goals down into objectives, then design specific tactics and initiatives to move toward these objectives. Each tactic or initiative must have one or more specific expected outcomes, measured in terms of the concrete measures of success. Ultimately, the sum of the expected outcomes should embody the strategic goals.

Conclusions and Limitations

It is possible for school superintendents and school principals, based on this work, with a degree of will and perseverance, to implement a strategic transformation in a public elementary school using the Climate4Creativity measures and the definitions embedded in the Climate4Creativity constructs, as part of a management system based on theory of constraints.

By exploring school safety, community, and the structure of the learning environment required for the teaching and learning of higher order thinking in a public elementary school, this work begins the creation of a framework to enable school leaders to make significant, transformational, strategic change in their schools. This work provides a small piece of the solution to the problem of modern school improvement and helps move society toward a more sustainable education system. The Climate4Creativity opportunity cycle (Figure 39), illustrates how this work fits into the wider solution space.

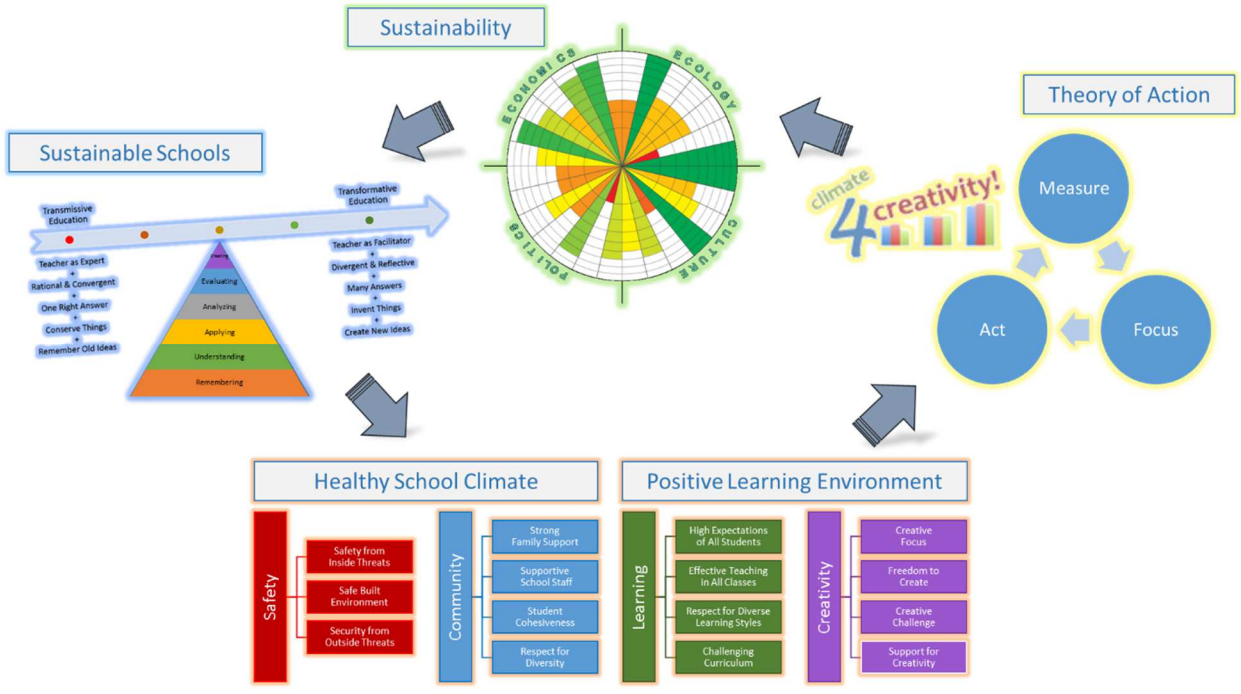


Figure 39. The Climate4Creativity Opportunity Cycle.

Limitations. The first limitation of the work completed here is the student developmental breadth of the instruments – to fully realize the early promise of this work, an early childhood (grades PK-2) version of the instrument and a high school (grades 9-12) version of the instrument should be developed and tested alongside the elementary (grades 3-8) instruments. This could potentially be extended into undergraduate college communities, with some adaptation. Each instrument should be linked to the higher and lower grade level instruments through a series of formal equating studies, allowing analysis of developmental shifts between instruments as well as within each instrument.

Second, family perspectives and staff perspectives instruments should be developed, perhaps along with a community perspectives instrument, to provide school leaders with a more complete, 360º landscape of their school. The addition of these perspectives would provide additional support, not only for understanding the constraints on the system, but also for creation of the guiding coalition to help improve the school.

Third, an action research project should be conducted, to use the fully developed theory of action in a real school improvement project over a period of 3-5 years. This work should be focused on developing management decision making protocols and methods, on the integration of academic outcome measures (such as standardized test results) along with the Climate4Creativity measures, and on completing other validation work (such as using observational methods to assess the learning environment, in order to create a multi-method multi-trait matrix). The theory of Climate4Creativity predicts that improvement in the safety, community, and learning environment in a public elementary school would lead to improved academic outcomes for a school overall, both in terms of standardized tests and broader non-standardized outcomes.

Finally, an action planning guide, expanded measures (both grade levels and additional perspectives), and updated construct design, along with evidence-based findings from the research should be integrated into a *Climate4Creativity School Improvement Kit*, which could be made available for school leaders as both a training workshop and supported program of action research for interested schools. This guide should encompass generic school improvement strategies, and describe how the instruments can be used at a detailed item level and at a measure level to select and refine an appropriate strategies, and ultimately, to manage execution of an improvement plan in a particular school. This has the potential to become a useful *how to* guide for school leaders, supported by research, and tools for school leaders to take action to improve outcomes for students in their schools.

A Call to Action

Significant, transformational, sustainable change is hard, particularly in complex institutional and organizational environments. By enhancing the practice of school improvement with a specific focus on *teaching and learning creativity*, an opportunity emerges to better prepare young people to participate in the social life of the planet, thereby empowering the sustainability of human society. To achieve this, an action-oriented *theory of climate for creativity* was proposed in chapter 4, a measure of the theory constructed in chapters 5 and 6, and the intended uses and interpretations of its scores were validated in chapter 7 and in this chapter.

Equipped with a theory of action and a measure of success, school leaders are enabled to focus and act to improve their schools. The Climate4Creativity project provides school leaders with the tools they need to improve their schools. This is an urgent call to action: Public school leaders across the U.S. need to study, embrace, inspire, lead, and manage transformational change in their schools.

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78*(4), 1102-1134. doi: 10.3102/0034654308326084

An Act Concerning the Strengthening of School Bullying Laws of 2011, Pub. L. No. 11-232 (Conn. Gen. Stat. 2011).

Adamski, A., Fraser, B. J., & Peiro, M. M. (2013). Parental involvement in schooling, classroom environment and student outcomes. *Learning Environments Research, 16*(3), 315-328. doi: 10.1007/s10984-012-9121-7

Agogué, M., Poirel, N., Pineau, A., Houdé, O., & Cassotti, M. (2014). The impact of age and training on creativity: A design-theory approach to study fixation effects. *Thinking Skills and Creativity, 11*, 33-41. doi: 10.1016/j.tsc.2013.10.002

Amabile, T. M., Hadley, C. N., & Kramer, S. J. (2002). Creativity under the gun. *Harvard Business Review, August 2002*.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Addison Wesley Longman.

Antes, A. L., & Mumford, M. D. (2009). Effects of time frame on creative thought: Process versus problem-solving effects. *Creativity Research Journal, 21*(2-3), 166-182. doi: 10.1080/10400410902855267

Astor, R. A., Benbenishty, R., & Estrada, J. N. (2009). School violence and theoretically atypical schools: The principal's centrality in orchestrating safe schools. *American Educational Research Journal, 46*(2), 423-461. doi: 10.3102/0002831208329598

Astor, R. A., Guerra, N., & Van Acker, R. (2010). How can we improve school safety research? *Educational Researcher, 39*(1), 69-78. doi: 10.3102/0013189x09357619

Astor, R. A., & Meyer, H. A. (2001). The conceptualization of violence-prone school subcontexts: Is the sum of the parts greater than the whole? *Urban Education, 36*, 374-399.

Astor, R. A., Meyer, H. A., & Pitner, R. O. (2001). Elementary and middle school students' perceptions of violence-prone school subcontexts. *The Elementary School Journal, 101*, 511-528.

Batanova, M., Espelage, D. L., & Rao, M. A. (2014). Early adolescents' willingness to intervene: What roles do attributions, affect, coping, and self-reported victimization play? *Journal of School Psychology*. doi: 10.1016/j.jsp.2014.02.001

Beer, M., & Nohria, N. (2000). Cracking the code of change. *Harvard Business Review, May-June 2000*.

Beghetto, R. A., & Kaufman, J. C. (2009). Intellectual estuaries: Connecting learning and creativity in programs of advanced academics. *Journal of Advanced Academics, 20*(2), 296-324.

Bergsmann, E. M., Van De Schoot, R., Schober, B., Finsterwald, M., & Spiel, C. (2013). The effect of classroom structure on verbal and physical aggression among peers: a short-term longitudinal study. *Journal of School Psychology, 51*(2), 159-174. doi: 10.1016/j.jsp.2012.10.003

Bishop, R., Ladwig, J., & Berryman, M. (2013). The centrality of relationships for pedagogy: The Whanaungatanga thesis. *American Educational Research Journal, 51*(1), 184-214. doi: 10.3102/0002831213510019

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive Domain*. White Plains, NY: Longman.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). New York, NY: Routledge.

Booth, A., Papaopannou, D., & Sutton, A. (2012). *Systematic approaches to a successful literature review*. Thousand Oaks, CA: Sage.

Borum, R., Cornell, D. G., Modzeleski, W., & Jimerson, S. R. (2010). What can be done about school shootings?: A review of the evidence. *Educational Researcher, 39*(1), 27-37. doi: 10.3102/0013189x09357620

Bouman, T., van der Meulen, M., Goossens, F. A., Olthof, T., Vermande, M. M., & Aleva, E. A. (2012). Peer and self-reports of victimization and bullying: their differential association with internalizing problems and social adjustment. *Journal of School Psychology, 50*(6), 759-774. doi: 10.1016/j.jsp.2012.08.004

Brock, L. L., Nishida, T. K., Chiong, C., Grimm, K. J., & Rimm-Kaufman, S. E. (2008). Children's perceptions of the classroom environment and social and academic performance: a longitudinal analysis of the contribution of the Responsive Classroom approach. *Journal of School Psychology, 46*(2), 129-149. doi: 10.1016/j.jsp.2007.02.004

Chan, C. S., Rhodes, J. E., Howard, W. J., Lowe, S. R., Schwartz, S. E., & Herrera, C. (2013). Pathways of influence in school-based mentoring: the mediating role of parent and teacher relationships. *Journal of School Psychology, 51*(1), 129-142. doi: 10.1016/j.jsp.2012.10.001

Chapman, R. L., Buckley, L., Sheehan, M. C., Shochet, I. M., & Romaniuk, M. (2011). The impact of school connectedness on violent behavior, transport risk-taking behavior, and associated injuries in adolescence. *Journal of School Psychology, 49*(4), 399-410. doi: 10.1016/j.jsp.2011.04.004

Cheng, V. M. Y. (2011). Infusing creativity into Eastern classrooms: Evaluations from student perspectives. *Thinking Skills and Creativity, 6*(1), 67-87. doi: 10.1016/j.tsc.2010.05.001

Chiu, M. M. (2008). Effects of argumentation on group micro-creativity: Statistical discourse analyses of algebra students' collaborative problem solving. *Contemporary Educational Psychology, 33*(3), 382-402. doi: 10.1016/j.cedpsych.2008.05.001

Chung, T.-s. (2013). Table-top role playing game and creativity. *Thinking Skills and Creativity, 8*, 56-71. doi: 10.1016/j.tsc.2012.06.002

Ciani, K. D., Middleton, M. J., Summers, J. J., & Sheldon, K. M. (2010). Buffering against performance classroom goal structures: The importance of autonomy support and classroom community. *Contemporary Educational Psychology, 35*(1), 88-99. doi: 10.1016/j.cedpsych.2009.11.001

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.): Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159. doi: 10.1037/0033-2909.112.1.155

Colvin, G., Sugai, G., Good, R. H., & Lee, Y. (1997). Using active supervision and precorrection to improve transition behaviors in an elementary school. *School Psychology Quarterly, 12*, 344-363.

Common Core Standards Initiative (Producer). (2012). Mission statement. *Common Core Standards Initiative*. Retrieved from http://www.corestandards.org

Connecticut Coalition for Justice in Education Funding, Inc., v. Governor M. Jodi Rell et al. (Supreme Court of Connecticut 2010).

Connecticut State Department of Education. (2014). Connecticut Education Data and Research. Retrieved August 4, 2014, from http://sdeportal.ct.gov/Cedar/WEB/ct_report/CedarHome.aspx?

Connor, C. M., Morrison, F. J., Fishman, B. J., Ponitz, C. C., Glasney, S., Underwood, P. S., . . . Schatschneider, C. (2009). The ISI Classroom Observation System: Examining the literacy instruction provided to individual students. *Educational Researcher, 38*(2), 85-99. doi: 10.3102/0013189x09332373

Connor, C. M., Spencer, M., Day, S. L., Giuliani, S., Ingebrand, S. W., McLean, L., & Morrison, F. J. (2014). Capturing the complexity: Content, type, and amount of instruction and quality of the classroom learning environment synergistically predict third graders' vocabulary and reading comprehension outcomes. *Journal of Educational Psychology*. doi: 10.1037/a0035921

Cook, T. D., & Hirschfield, P. J. (2008). Comer's School Development Program in Chicago: Effects on involvement with the Juvenile Justice System from the late elementary through the high school years. *American Educational Research Journal, 45*(1), 38-67. doi: 10.3102/0002831207308648

Cooper, K. S. (2013). Eliciting engagement in the high school classroom: A mixed-methods examination of teaching practices. *American Educational Research Journal, 51*(2), 363-402. doi: 10.3102/0002831213507973

Cornell, D. G., Gregory, A., Huang, F., & Fan, X. (2013). Perceived prevalence of teasing and bullying predicts high school dropout rates. *Journal of Educational Psychology, 105*(1), 138-149. doi: 10.1037/a0030416

Cornell, D. G., & Williams, F. (2012). Student threat assessment as a strategy to reduce school violence. In S. R. Jimerson, A. B. Nickerson, M. J. Mayer, & M. J. Furlong (Eds.), *Handbook of School Violence and School Safety: International Research and Practice* (Second ed., pp. 503-514). New York, NY: Routledge.

Corpus, J. H., McClintic-Gilbert, M. S., & Hayenga, A. O. (2009). Within-year changes in children's intrinsic and extrinsic motivational orientations: Contextual predictors and academic outcomes. *Contemporary Educational Psychology, 34*(2), 154-166. doi: 10.1016/j.cedpsych.2009.01.001

Cronbach, L. J. (1971). Test Validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. Thousand Oaks, CA: Sage.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety: Experiencing flow in work and play*. San Francisco, CA: Jossey-Bass.

Csikszentmihalyi, M. (1988). Society, culture, snd person: A system view of creativity. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 325-339). New York, NY: Cambridge University Press.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: HarperPerennial.

Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention*. New York, NY: HarperPerennial.

Curby, T. W., Rimm-Kaufman, S. E., & Abry, T. (2013). Do emotional support and classroom organization earlier in the year set the stage for higher quality instruction? *Journal of School Psychology, 51*(5), 557-569. doi: 10.1016/j.jsp.2013.06.001

Danielsen, A. G., Wiium, N., Wilhelmsen, B. U., & Wold, B. (2010). Perceived support provided by teachers and classmates and students' self-reported academic initiative. *Journal of School Psychology, 48*(3), 247-267. doi: 10.1016/j.jsp.2010.02.002

Davies, D., Jindal-Snape, D., Collier, C., Digby, R., Hay, P., & Howe, A. (2013). Creative learning environments in education—A systematic literature review. *Thinking Skills and Creativity, 8*, 80-91. doi: 10.1016/j.tsc.2012.07.004

Dewey, J. (1938). *Experience and education*. New York, NY: Touchstone.

DiLiello, T. C., Houghton, J. D., & Dawley, D. (2011). Narrowing the creativity gap: The moderating effects of perceived support for creativity. *Journal of Psychology, 145*(3), 151-172. doi: 10.1080/00223980.2010.548412

Duck, J. D. (1993). Managing change: The art of balancing. *Harvard Business Review, Nov-Dec 1993*.

Dukes, R. L., Stein, J. A., & Zane, J. I. (2010). Gender differences in the relative impact of physical and relational bullying on adolescent injury and weapon carrying. *Journal of School Psychology, 48*(6), 511-532. doi: 10.1016/j.jsp.2010.08.001

Eliot, M., Cornell, D. G., Gregory, A., & Fan, X. (2010). Supportive school climate and student willingness to seek help for bullying and threats of violence. *Journal of School Psychology, 48*(6), 533-553. doi: 10.1016/j.jsp.2010.07.001

Estell, D. B., Farmer, T. W., Pearl, R., Van Acker, R., & Rodkin, P. C. (2008). Social status and aggressive and disruptive behavior in girls: individual, group, and classroom influences. *Journal of School Psychology, 46*(2), 193-212. doi: 10.1016/j.jsp.2007.03.004

Eyman, W., & Cohen, J. (2009). Breaking the bully-victim-bystander cycle tool kit. New York, NY: National School Climate Center.

Frey, K. S., Hirschstein, M. K., Edstrom, L. V., & Snell, J. L. (2009). Observed reductions in school bullying, nonbullying aggression, and destructive bystander behavior: A longitudinal evaluation. *Journal of Educational Psychology, 101*(2), 466-481. doi: 10.1037/a0013839

Gentry, M. (2006). No Child Left Behind: Neglecting excellence. *Roeper Review 29*, 24-27.

Gini, G., Pozzoli, T., Borghi, F., & Franzoni, L. (2008). The role of bystanders in students' perception of bullying and sense of safety. *Journal of School Psychology, 46*(6), 617-638. doi: 10.1016/j.jsp.2008.02.001

Givens Rolland, R. (2012). Synthesizing the evidence on classroom goal structures in middle and secondary schools: A meta-analysis and narrative review. *Review of Educational Research, 82*(4), 396-435. doi: 10.3102/0034654312464909

Goetz, T., Lüdtke, O., Nett, U. E., Keller, M. M., & Lipnevich, A. A. (2013). Characteristics of teaching and students' emotions in the classroom: Investigating differences across domains. *Contemporary Educational Psychology, 38*(4), 383-394. doi: 10.1016/j.cedpsych.2013.08.001

Goldratt, E. M. (2010). What is TOC? In J. F. Cox III & J. G. Schleier (Eds.), *Theory of Constraints Handbook* (pp. 3-12). New York, NY: McGraw-Hill.

Goldratt, E. M., & Cox, J. (1984, 2004). *The goal: A process of ongoing improvement* (3rd ed.). Great Barrington, MA: North River Press.

Goldsmith, P. R. (2010). Coleman revisited: School segregation, peers, and frog ponds. *American Educational Research Journal, 48*(3), 508-535. doi: 10.3102/0002831210392019

Goldweber, A., Waasdorp, T. E., & Bradshaw, C. P. (2013). Examining the link between forms of bullying behaviors and perceptions of safety and belonging among secondary school students. *Journal of School Psychology, 51*(4), 469-485. doi: 10.1016/j.jsp.2013.04.004

Gregory, A., Cornell, D. G., & Fan, X. (2011). The relationship of school structure and support to suspension rates for black and white high school students. *American Educational Research Journal, 48*(4), 904-934. doi: 10.3102/0002831211398531

Gregory, A., Cornell, D. G., Fan, X., Sheras, P., Shih, T.-H., & Huang, F. (2010). Authoritative school discipline: High school practices associated with lower bullying and victimization. *Journal of Educational Psychology, 102*(2), 483-496. doi: 10.1037/a0018562

Guerrero-Bote, V. P., & Moya-Anegón, F. (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of Informetrics, 6*(4), 674-688. doi: 10.1016/j.joi.2012.07.001

Hand, V. M. (2009). The co-construction of opposition in a low-track mathematics classroom. *American Educational Research Journal, 47*(1), 97-132. doi: 10.3102/0002831209344216

Hemphill, F. C., & Vanneman, A. (2011). Achievement gaps: How Hispanic and white students in public schools perform in mathematics and reading on the National Assessment of Educational Progress (NCES 2011-459). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Henry, D. B., Farrell, A. D., Schoeny, M. E., Tolan, P. H., & Dymnicki, A. B. (2011). Influence of school-level variables on aggression and associated attitudes of middle school students. *Journal of School Psychology, 49*(5), 481-503. doi: 10.1016/j.jsp.2011.04.007

Ho, H.-C., Wang, C.-C., & Cheng, Y.-Y. (2013). Analysis of the scientific imagination process. *Thinking Skills and Creativity, 10*, 68-78. doi: 10.1016/j.tsc.2013.04.003

Hoover, J. H., Oliver, R. L., & Thomson, K. A. (1993). Perceived victimization by school bullies: New research and future direction. *Journal of Humanistic Education and Development, 32*, 76-84.

Huxley, A. (1932). *Brave new world*. London, UK: HarperCollins.

Huxley, A. (1958). *Brave new world revisited*. London, UK: HarperCollins.

Isaksen, S. G., & Isaksen, E. J. (2010). The climate for creativity and innovation: And its relationship to empowerment, customer insight and ambiguity. New York, NY: Creative Problem Solving Group, Inc.

James, P. (2015). *Urban sustainability in theory and practice: Circles of sustainability*. New York, NY: Routledge.

Janssen, I., Craig, W. M., Boyce, W. F., & Pickett, W. (2004). Associations between overweight and obesity with bullying behaviors in school-aged children. *Pediatrics, 113*, 1187-1194.

Jefferson, T. (1816). Letter from Thomas Jefferson to Charles Yancey *Quotes by and about Thomas Jefferson*. Washington, DC: Library of Congress.

Jimerson, S. R., Nickerson, A. B., Mayer, M. J., & Furlong, M. J. (Eds.). (2012). *Handbook of school violence and school safety: International research and practice* (2nd ed.). New York, NY: Routledge.

Jutengren, G., Kerr, M., & Stattin, H. (2011). Adolescents' deliberate self-harm, interpersonal stress, and the moderating effects of self-regulation: a two-wave longitudinal analysis. *Journal of School Psychology, 49*(2), 249-264. doi: 10.1016/j.jsp.2010.11.001

Juvonen, J., Graham, S., & Schuster, M. A. (2003). Bullying among young adolescents: The strong, the weak, and the troubled. *Pediatrics, 112*, 1231-1237.

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). New York, NY: American Council on Education/Praeger.

Kangas, M. (2010). Finnish children's views on the ideal school and learning environment. *Learning Environments Research, 13*(3), 205-223. doi: 10.1007/s10984-010-9075-6

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four C model of creativity. *Review of General Psychology, 13*(1), 1-12.

Kaufman, J. C., & Sternberg, R. J. (2007). Resource review: Creativity. *Change 39*, 55-58.

Kaufman, J. C., & Sternberg, R. J. (2010). *The Cambridge handbook of creativity*. New York, NY: Cambridge University Press.

Kim, K. H. (2008). Underachievement and creativity: Are gifted underachievers highly creative? *Creativity Research Journal, 20*(2), 234-242. doi: 10.1080/10400410802060232

Kim, K. H. (2011). The creativity crisis: The decrease in creative thinking scores on the Torrance Tests of Creative Thinking. *Creativity Research Journal*.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Upper Saddle River, NJ: Prentice Hall.

Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2008). A multilevel study of predictors of student perceptions of school climate: The effect of classroom-level factors. *Journal of Educational Psychology, 100*(1), 96-104. doi: 10.1037/0022-0663.100.1.96

Kreitzer, A. E., & Madaus, G. F. (1994). Empirical investigations of the hierarchical structure of the taxonomy. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's taxonomy: A forty-year retrospective, Ninety-third Yearbook of the National Society for the Study of Education* (pp. 64-81). Chicago, IL: Chicago University Press.

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. New York, NY: Penguin.

Lassig, C. J. (2013). Approaches to creativity: How adolescents engage in the creative process. *Thinking Skills and Creativity, 10*, 3-12. doi: 10.1016/j.tsc.2013.05.002

Lawson, M. A., & Alameda-Lawson, T. (2011). A case study of school-linked, collective parent engagement. *American Educational Research Journal, 49*(4), 651-684. doi: 10.3102/0002831211427206

Leach, L. P. (2004). *Critical chain project management* (2nd ed.). Norwood, MA: Artech House.

Lee, J. (2010). Tripartite growth trajectories of reading and math achievement: Tracking national academic progress at primary, middle, and high school levels. *American Educational Research Journal, 47*(4), 800-832. doi: 10.3102/0002831210365009

Leung, A. K. Y., & Chiu, C.-y. (2008). Interactive effects of multicultural experiences and openness to experience on creative potential. *Creativity Research Journal, 20*(4), 376-382. doi: 10.1080/10400410802391371

Linacre, J. M. (2012). *Winsteps(R) Rasch measurement computer program user's guide*. Beaverton, OR: Winsteps.com.

Linacre, J. M. (2014). Winsteps (Version 3.81.0). Beaverton, OR: Winsteps.com.

Little, L. (2002). Middle-class mothers' perceptions of peer and sibling victimization among children with Asperger's syndrome and nonverbal learning disorders. *Issues in Comprehensive Pediatric Nursing, 25*, 43-57.

Long, H. (2014). More than appropriateness and novelty: Judges' criteria of assessing creative products in science tasks. *Thinking Skills and Creativity, 13*, 183-194. doi: 10.1016/j.tsc.2014.05.002

Loukas, A., Ripperger-Suhler, K. G., & Herrera, D. E. (2012). Examining competing models of the associations among peer victimization, adjustment problems, and school connectedness. *Journal of School Psychology, 50*(6), 825-840. doi: 10.1016/j.jsp.2012.07.003

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34*(2), 120-131. doi: 10.1016/j.cedpsych.2008.12.001

Mainemelis, C. (2002). Time and timelessness: Creativity in (and out of) the temporal dimension. *Creativity Research Journal, 14*(2), 227-238. doi: 10.1207/s15326934crj1402_9

Mainhard, M. T., Brekelmans, M., Brok, P. d., & Wubbels, T. (2011). The development of the classroom social climate during the first months of the school year. *Contemporary Educational Psychology, 36*(3), 190-200. doi: 10.1016/j.cedpsych.2010.06.002

Mann, H. (1848). Twelfth annual report to the secretary of the Massachusetts state board of education. Boston, MA: Massachusetts State Board of Education.

Mantel, N. (1963). Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association, 58*, 690-700.

Marini, Z., Fairbairn, L., & Zuber, R. (2001). Peer harassment in individuals with developmental disabilities: Towards the development of a multidimensional bullying identification model. *Development Disabilities Bulletin, 29*, 170-195.

Martin, A. J. (2008). Enhancing student motivation and engagement: The effects of a multidimensional intervention. *Contemporary Educational Psychology, 33*(2), 239-269. doi: 10.1016/j.cedpsych.2006.11.003

Martin, A. J., & Dowson, M. (2009). Interpersonal relationships, motivation, engagement, and achievement: Yields for theory, current issues, and educational practice. *Review of Educational Research, 79*(1), 327-365. doi: 10.3102/0034654308325583

Marzano, R. J., Marzano, J. S., & Pickering, D. J. (2003). *Classroom management that works: research-based strategies for every teacher*. Alexandria, VA: Association for Supervision and Curriculum Development.

Mathews, S. R., & Lowe, K. (2011). Classroom environments that foster a disposition for critical thinking. *Learning Environments Research, 14*(1), 59-73. doi: 10.1007/s10984-011-9082-2

Mayer, M. J., & Furlong, M. J. (2010). How safe are our schools? *Educational Researcher, 39*(1), 16-26. doi: 10.3102/0013189x09357617

McCurdy, B. L., Lannie, A. L., & Barnabas, E. (2009). Reducing disruptive behavior in an urban school cafeteria: An extension of the Good Behavior Game. *Journal of School Psychology, 47*(1), 39-54. doi: 10.1016/j.jsp.2008.09.003

McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology, 46*(3), 235-261. doi: 10.1016/j.jsp.2007.05.001

McLellan, R., & Nicholl, B. (2013). Creativity in crisis in design & technology: Are classroom climates conducive for creativity in English secondary schools? *Thinking Skills and Creativity, 9*, 165-185. doi: 10.1016/j.tsc.2012.11.004

Meadows, D. H. (2008). *Thinking in systems: A primer*. White River Junction, VT: Chelsea Green.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: American Council on Education/Macmillan.

Microsoft. (2013a). Excel (Version 2013). Redmond, WA: Microsoft.

Microsoft. (2013b). PowerPoint (Version 2013). Redmond, WA: Microsoft.

Microsoft. (2013c). Visio (Version 2013). Redmond, WA: Microsoft.

Microsoft. (2013d). Word (Version 2013). Redmond, WA: Microsoft.

Miller, D. N. (2012). Youth suicidal behavior in the context of school violence. In S. R. Jimerson, A. B. Nickerson, M. J. Mayer, & M. J. Furlong (Eds.), *Handbook of School Violence and School Safety: International Research and Practice* (Second ed., pp. 203-214). New York, NY: Routledge.

Mitchell, M. M., & Bradshaw, C. P. (2013). Examining classroom influences on student perceptions of school climate: the role of classroom management and exclusionary discipline strategies. *Journal of School Psychology, 51*(5), 599-610. doi: 10.1016/j.jsp.2013.05.005

Mondale, S., & Patton, S. (2001). *School: The story of American public education.* Boston, MA: Beacon Press.

Monsen, J. J., Ewing, D. L., & Kwoka, M. (2013). Teachers' attitudes towards inclusion, perceived adequacy of support and classroom learning environment. *Learning Environments Research, 17*(1), 113-126. doi: 10.1007/s10984-013-9144-8

Nansel, T. R., Haynie, D. L., & Simons-Morton, B. G. (2003). The association of bullying and victimization with middle school adjustment. *Journal of Applied School Psychology, 19*, 45-61.

National School Climate Council. (2007). *The school climate challenge: Narrowing the gap between school climate research and school climate policy, practice guidelines and teacher education policy*. New York, NY: National School Climate Council Retrieved from http://www.schoolclimate.org/climate/documents/school-climate-challenge.pdf.

Newton, D. P. (2013). Moods, emotions and creative thinking: A framework for teaching. *Thinking Skills and Creativity, 8*, 34-44. doi: 10.1016/j.tsc.2012.05.006

Nickerson, A. B., Mele, D., & Princiotta, D. (2008). Attachment and empathy as predictors of roles as defenders or outsiders in bullying interactions. *Journal of School Psychology, 46*(6), 687-703. doi: 10.1016/j.jsp.2008.06.002

Niehaus, K., Rudasill, K. M., & Rakes, C. R. (2012). A longitudinal study of school connectedness and academic outcomes across sixth grade. *Journal of School Psychology, 50*(4), 443-460. doi: 10.1016/j.jsp.2012.03.002

Norwich, B., & Kelly, N. (2004). Pupils' views on inclusion: Moderate learning difficulties and bullying in mainstream and special schools. *British Educational Research Journal, 30*, 43-65.

O'Connor, E. E. (2010). Teacher-child relationships as dynamic systems. *Journal of School Psychology, 48*(3), 187-218. doi: 10.1016/j.jsp.2010.01.001

O'Connor, E. E., Dearing, E., & Collins, B. A. (2010). Teacher-child relationship and behavior problem trajectories in elementary school. *American Educational Research Journal, 48*(1), 120-162. doi: 10.3102/0002831210365008

Olthof, T., Goossens, F. A., Vermande, M. M., Aleva, E. A., & van der Meulen, M. (2011). Bullying as strategic behavior: relations with desired and acquired dominance in the peer group. *Journal of School Psychology, 49*(3), 339-359. doi: 10.1016/j.jsp.2011.03.003

Olweus, D. (1995). Bullying or peer abuse at school: Facts and interventions. *Current Directions in Psychological Science, 4*, 196-200.

Orwell, G. (1949). *Nineteen Eighty-Four*. London, UK: Secker & Warburg.

Osher, D., Bear, G. G., Sprague, J. R., & Doyle, W. (2010). How can we improve school discipline? *Educational Researcher, 39*(1), 48-58. doi: 10.3102/0013189x09357618

O'Toole, M. E. (2000). The school shooter: A threat assessment perspective. Quantico, VA: National Center for the Analysis of Violent Crime, Federal Bureau of Investigation.

Patall, E. A., Cooper, H., & Robinson, J. C. (2008). Parent involvement in homework: A research synthesis. *Review of Educational Research, 78*(4), 1039-1101. doi: 10.3102/0034654308325185

Peng, S.-L., Cherng, B.-L., Chen, H.-C., & Lin, Y.-Y. (2013). A model of contextual and personal motivations in creativity: How do the classroom goal structures influence creativity via self-determination motivations? *Thinking Skills and Creativity, 10*, 50-67. doi: 10.1016/j.tsc.2013.06.004

Peter, F., & Dalbert, C. (2010). Do my teachers treat me justly? Implications of students' justice experience for class climate experience. *Contemporary Educational Psychology, 35*(4), 297-305. doi: 10.1016/j.cedpsych.2010.06.001

Plank, S. B., & Condliffe, B. F. (2013). Pressures of the season: An examination of classroom quality and high-stakes accountability. *American Educational Research Journal, 50*(5), 1152-1182. doi: 10.3102/0002831213500691

Postman, N. (1985). *Amusing ourselves to death: Public discourse in the age of show business*. London, UK: Penguin.

Pryce, S., & Frederickson, N. (2013). Bullying behaviour, intentions and classroom ecology. *Learning Environments Research, 16*(2), 183-199. doi: 10.1007/s10984-013-9137-7

QSR. (2012). NVivo (Version 10). Melbourne, Australia: QSR International Pty. Ltd.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedogogiske Institut.

Ravitsh, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education.* . New York, NY: Basic Books.

Reis, S. M., McCoach, D. B., Little, C. A., Muller, L. M., & Kaniskan, R. B. (2010). The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. *American Educational Research Journal, 48*(2), 462-501. doi: 10.3102/0002831210382891

Renzulli, J. S. (1992). A general theory for the development of creative productivity through the pursuit of ideal acts of learning. *Gifted Child Quarterly, 36*, 170-182.

Reznitskaya, A., Glina, M., Carolan, B., Michaud, O., Rogers, J., & Sequeira, L. (2012). Examining transfer effects from dialogic discussions to new tasks and contexts. *Contemporary Educational Psychology, 37*(4), 288-306. doi: 10.1016/j.cedpsych.2012.02.003

Rimm-Kaufman, S. E., Larsen, R. A. A., Baroody, A. E., Curby, T. W., Ko, M., Thomas, J. B., . . . DeCoster, J. (2014). Efficacy of the responsive classroom approach: Results From a 3-year, longitudinal randomized controlled trial. *American Educational Research Journal, 51*(3), 567-603. doi: 10.3102/0002831214523821

Robinson, J. P., & Espelage, D. L. (2011). Inequities in educational and psychological outcomes between LGBTQ and straight students in middle and high school. *Educational Researcher, 40*(7), 315-330. doi: 10.3102/0013189x11422112

Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research*. doi: 10.3102/0034654311421793

Rubenstein, L. D., McCoach, D. B., & Siegle, D. (2013). Teaching for creativity scales: An instrument to examine teachers' perceptions of factors that allow for the teaching of creativity. *Creativity Research Journal, 25*(3), 324-334. doi: 10.1080/10400419.2013.813807

Rubie-Davies, C. M., Peterson, E. R., Sibley, C. G., & Rosenthal, R. (2014). A teacher expectation intervention: Modelling the practices of high expectation teachers. *Contemporary Educational Psychology*. doi: 10.1016/j.cedpsych.2014.03.003

Rudasill, K. M., Reio, T. G., Jr., Stipanovic, N., & Taylor, J. E. (2010). A longitudinal study of student-teacher relationship quality, difficult temperament, and risky behavior from childhood to early adolescence. *Journal of School Psychology, 48*(5), 389-412. doi: 10.1016/j.jsp.2010.05.001

Runco, M. A. (2007). *Creativity theories and themes: Research, development, and practice*. Burlington, MA: Elsevier Academic Press.

Sawyer, R. K. (2006). *Explaining creativity: The science of human innovation*. New York, NY: Oxford University Press.

Scherr, T. G. (2012). Addressing the needs of marginalized youth at school. In S. R. Jimerson, A. B. Nickerson, M. J. Mayer, & M. J. Furlong (Eds.), *Handbook of School Violence and*

School Safety: International Research and Practice (Second ed., pp. 105-115). New York, NY: Routledge.

SCImago. (2007). SCImago journal & country rankings. Retrieved January 14, 2014, from http://www.scimagojr.com

Shane, H. G. (1981). Significant writings that have influenced the curriculum: 1906-81. *Phi Delta Kappan, 63*, 311-314.

Spearman, J., & Watt, H. M. G. (2013). Perception shapes experience: The influence of actual and perceived classroom environment dimensions on girls' motivations for science. *Learning Environments Research, 16*(2), 217-238. doi: 10.1007/s10984-013-9129-7

Sterling, S. (2001). *Sustainable education: Re-visioning learning and change*. Devon, England: Green Books.

Sullivan, M. (Producer). (2011). Behind America's tutor boom. *Wall Street Journal MarketWatch Report*. Retrieved from http://www.marketwatch.com/story/behind-americas-tutor-boom-1318016970246

Swearer, S. M., Espelage, D. L., Vaillancourt, T., & Hymel, S. (2010). What can be done about school bullying?: Linking research to educational practice. *Educational Researcher, 39*(1), 38-47. doi: 10.3102/0013189x09357622

Syvertsen, A. K., Flanagan, C. A., & Stout, M. D. (2009). Code of silence: Students' perceptions of school climate and willingness to intervene in a peer's dangerous plan. *Journal of Educational Psychology, 101*(1), 219. doi: 10.1037/a0013246

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research, 83*(3), 357-385. doi: 10.3102/0034654313483907

ThomsonReuters. (2013). EndNote (Version X7 for Windows). New York, NY: Thomson Reuters.

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. San Francisco, CA: Jossey-Bass.

Turner, J. C., Gray, D. L., Anderman, L. H., Dawson, H. S., & Anderman, E. M. (2013). Getting to know my teacher: Does the relation between perceived mastery goal structures and perceived teacher support change across the school year? *Contemporary Educational Psychology, 38*(4), 316-327. doi: 10.1016/j.cedpsych.2013.06.003

Veenstra, R., Lindenberg, S., Huitsing, G., Sainio, M., & Salmivalli, C. (2014). The role of teachers in bullying: The relation between antibullying attitudes, efficacy, and efforts to reduce bullying. *Journal of Educational Psychology*. doi: 10.1037/a0036110

Vossekuil, B., Fein, R. A., Reddy, M., Borum, R., & Modzeleski, W. (2002). The final report and findings of the safe school initiative: Implications for the prevention of school attacks in the United States. Washington, DC: United States Secret Service and United States Department of Education.

Waasdorp, T. E., & Bradshaw, C. P. (2011). Examining student responses to frequent bullying: A latent class approach. *Journal of Educational Psychology, 103*(2), 336-352. doi: 10.1037/a0022747

Wallach, M. A., & Kogan, N. (1965). A new look at creativity-intelligence distinction. *Journal of Personality, 33*, 348-369.

Wallas, G. (1926). *The art of thought*. New York, NY: Harcourt, Brace.

Wang, J., Iannotti, R. J., & Luk, J. W. (2012). Patterns of adolescent bullying behaviors: physical, verbal, exclusion, rumor, and cyber. *Journal of School Psychology, 50*(4), 521-534. doi: 10.1016/j.jsp.2012.03.004

Wang, M.-T., & Holcombe, R. (2010). Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Research Journal, 47*(3), 633-662. doi: 10.3102/0002831209361209

Williams, W. M., & Yang, L. T. (2010). Organizational creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity*. New York, NY: Cambridge University Press.

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: SAGE.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. New York, NY: Psychology Press.

Wirkala, C., & Kuhn, D. (2011). Problem-based learning in K-12 education: Is it effective and how does it achieve its effects? *American Educational Research Journal, 48*(5), 1157-1186. doi: 10.3102/0002831211419491

Woods, S., & Wolke, D. (2004). Direct and relational bullying among primary school children and academic achievement. *Journal of School Psychology, 42*, 135-155.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370.

Yale School of Medicine. (2009). Comer school development program. Retrieved 7/1/2012, from http://childstudycenter.yale.edu/comer/evaluation/surveys/scs/index.aspx

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337-347). Hillsdale, NJ: Lawrence Erlbaum.

Appendix A – Glossary of Terms

Table 32.
Glossary of Terms.

Term	Description
Accountability	A term used to categorize systems of public accountability in public school systems, where a variety of data is published to enable parents to understand the quality of schooling in their districts, and to make informed decisions about which schools their children should attend. In Connecticut, all school districts report their performance and discipline data annually to the state department of education, and the state publishes the data in searchable databases. *See also "parental choice."*
Andrich Threshold	The point of intersection between the probability curves of two categories in a polytomous measurement instrument, used to assess proper category functioning and other psychometric properties of a measure.
Autotelic Response, Automaticity	An autotelic response to a stimulus is a response which becomes automatic or instinctual, through repeated practice, common examples include writing or playing tennis, where the action requires little conscious thought. *See also "flow."*
Calibration	A comprehensive analytical process of mapping an instrument to its measures. Calibration provides evidence for validation of an instrument and prepares an initial benchmark against which future uses of the instrument can be compared.
Category Functioning	Categories in a multichoice item are expected to perform according to Rasch model parameters. Category functioning is an analytical technique to examine the degree to which categories actually do perform empirically according to model expectations.
Cognitive Lab, Cognitive Interview	A form of qualitative research where participants drawn from a target respondent population are interviewed about the items in an items design, in an attempt to explore their cognitive processes and thereby improve the quality of the items and answer choices.

Common Schools	A form of community schools in the early commonwealth of Massachusetts, championed by Mann, and generally seen as a precursor to US Public Schools.
Construct, Construct Map	A structured way of thinking about the idea or ideas being measured in an instrument. The construct IS the idea, and the construct map is a graphical method to describe the construct in a very precise and detailed way.
Differential Item Functioning (DIF)	A form of psychometric analysis which examines the subsamples of a respondent pool using contrasts such as gender, ethnicity, or age group in order to evaluate an instrument for bias based empirical performance of the instrument.
Experiential Education	A form of educational philosophy generally considered to have been founded by John Dewey, where students learn through authentic experiences in the world. This form of education is used extensively in outdoors and physical education, and has been applied in several systems of teacher practice including cooperative learning systems.
Fit Statistic	Fit statistics are measures of the degree to which an item or a complete items design fits the expectations of a Rasch measurement model. A mean square statistic is generally used, and rating scale (survey) instrument developers seek mean square values between 0.6 and 1.4 for reasonable fit. Mean square fit statistics above 2.0 are generally considered to degrade the quality of the measurement. Fit statistics between 1.5 and 2.0 are generally considered to add little meaningful information to the measure, but do not degrade the measure, fit statistics between 0.5 and 1.5 are considered to fit model expectations well, fit statistics below 0.5 are considered to fit the model too well, and therefore provide little meaningful information (Wright & Linacre, 1994). *See also "Rasch measurement," "infit statistic," and "outfit statistic."*
Heteroscedastic, heteroscedasticity	Heteroscedasticity is s property of the variability of data within a pair of samples or populations being analyze using statistical techniques. Heteroscedasticity implies that the variability of the two datasets is different, homoscedasticity implies that the variability of the two datasets is similar.

Homoscedastic, homoscedasticity	Homoscedasticity is s property of the variability of data within a pair of samples or populations being analyze using statistical techniques. Homoscedasticity implies that the variability of the two datasets is similar, Heteroscedasticity implies that the variability of the two datasets is different.
Infit Statistic	The infit statistic is a type of fit statistic focused on the core "inliers" within an empirical dataset. These are the data from the most information rich respondents. *See also "fit statistic."*
Instrument	A mechanism for measuring a thing or an attitude. For example, a ruler is an instrument designed to measure length.
Item	An item is a single question in a psychometric instrument design. Items can take the form of multichoice, multioption, freeform, or other item types.
Items Design	The items design is the collective item structure for an instrument, made up of a careful description of every item and all the answer choices for every item.
Latent Trait, Latent Attitude, Latent Knowledge	A latent trait is an underlying truth about a respondent, modeled by the respondent's choice of answer to an item or group of items in the items design.
Logit, Log Odds	A simple mathematical computation is used to convert raw scores expressed as probabilities into their logit equivalent. This has the net impact of stretching the raw scale, and reducing floor and ceiling effects. This mathematical technique is comparable to techniques using in the physical sciences when calculating derived measures.
Measurement Model	The measurement model is a logit model derived from the raw scores extracted from an instrument. There are a variety of measurement models in the Rasch Measurement model family. *See also "Rasch Measurement."*
Multi-trait Multi-method Matrix (MTMM)	A simple graphical approach to comparing a new instrument to other measures of the same or similar traits.
Outfit Statistic	The outfit statistic is a type of fit statistic focused on the outliers within an empirical dataset. These are the data

	from the least information rich, but potentially most error prone respondents. *See also "fit statistic."*
Parental Choice	Parental choice is a concept in modern education reform where parents are intended to be allowed to choose which school their children attend. It is often combined with accountability to imply that, equipped with data about school performance, parents can make better decisions about schooling. *See also "accountability."*
Pathway Analysis	An analytical technique in which empirical item data is compared to a "pathway" based on the measurement model being used. Items which do not "fit" are modified or removed from the instrument.
Psychometric Analysis	A branch of analysis which examines the technical qualities of a measurement instrument in the human sciences.
Rasch Measurement, Rasch Model	A family of measurement models developed from the original germinal work of Georg Rasch. Rasch models provide a person free, item free probabilistic model for measurement in the human sciences. The Rasch measurement model is a logit model derived from the raw scores extracted from an instrument. There are a variety of measurement models in the Rasch Measurement model family. *See also "Measurement Model."*
Raw Score	The category score extracted directly from an instrument (i.e. 1,2,3,4,5), which is then converted to a measurement score using logit units.
Reliability	Reliability is the concept of consistency in the operation of a psychometric instrument. It is typically expressed using a reliability statistic such as Cronbach's Alpha, although it can be shown that Cronbach's Alpha tends to overestimate reliability. In Rasch measurement applications, a more conservative error-adjusted person separation reliability statistic is often used in conjunction with Cronbach's Alpha, in order to provide a range of confidence of the reliability of a measure. *See also "Rasch measurement."*
SALSA Literature Review Methodology	A systematic literature review methodology – SALSA stands for search, appraisal, synthesis, analysis.

Scalogram	A method where items and persons are arranged in a matrix to reflect the item difficulty and person ability on a test. Items are ordered from easiest to hardest, and persons are ordered from highest ability to lowest.
Theory of Action	A theoretical path to action, defined in terms of desired outcomes, mapped to specific goals, mapped to specific actions.
Theory of Constraints	A systems thinking framework from operations research, focused on using the goals of a system, combined with a deep understanding of the constraints at any point in time to enable leaders to focus on what matters most in order to drive improvement into the system.
Think Aloud	A research technique designed to understand cognitive processes, where respondents are asked to think aloud about why they responded in the way they did.
Torrance Tests of Creative Thinking (TTCT)	The most widely used individual test of creative abilities.
Validation, Validity	Validation, in the context of this dissertation, is the concept of meaningfulness of the intended interpretations and uses of a measure. This dissertation uses a modern validity theory approach to validation.
Wright Map, Item-person Map	A Wright map displays the respondent "ability" distribution alongside the item "difficulty" distribution on a common logit scale, allowing the analyst to qualitatively examine the performance and range of a test.

Appendix B – Literature Review Methodology

The SALSA framework Booth et al. (2012) used in this study had five major phases, the first four of which were intended to prepare the researcher with documentation and insight to support the final writing phase of the literature review. The five phases are shown in Table 33. Each phase built progressively on the learnings from the prior phase, iteratively and systematically constructing a broad and deep understanding of the structure of the literature. A variety of software products were used in this review, including EndNote X7 (ThomsonReuters, 2013) and nVivo 10 for Windows (QSR, 2012), along with the Microsoft Excel (Microsoft, 2013a), Word (Microsoft, 2013d), PowerPoint (Microsoft, 2013b), and Visio (Microsoft, 2013c).

Table 33.
The SALSA Literature Review Framework.

	Search	**Appraisal**	**Synthesis**	**Analysis**	**Writing**
Description	Identify candidate articles	Appraise candidate articles	Identify themes within the literature	Analyze articles and themes	Write the literature review
Inputs	Journal rating data Journal access	Full text articles Appraisal rubric	Articles	Initial thematic structure Articles Germinal texts	Thematic structure Thematic maps Articles & texts Appraisal & lit review databases
Outputs	Selected journals Candidate articles	Selected articles Appraisal database	Initial thematic structure Synthesis diagrams	Final thematic structure Lit review database Thematic maps	Lit review chapter Figures & tables
Tools	Journal rankings Library databases Inter-library loan EndNote X7	EndNote X7 Excel 2013	EndNote X7 nVivo 10 Excel 2013 Visio 2013 PowerPoint 2013	EndNote X7 nVivo 10 Excel 2013 Visio 2013 PowerPoint 2013	EndNote X7 nVivo 10 Visio 2013 Excel 2013 Word 2013 PowerPoint 2013

Note. Adapted from Booth et al. (2012).

Conduct of the Review

First, a systematic search for articles was conducted using selected scholarly journals as sources. In this case, specific, carefully selected and relevant scholarly journals were used for the search. Iterative trial searches were conducted on these journals to develop an effective search screen, which was then used to identify a finite pool of candidate articles from each journal. These candidate articles were loaded into an article database. In the second, appraisal phase, the candidate articles identified in phase one were systematically evaluated using a predefined rubric, to select highly relevant, high quality articles for inclusion in the review. The appraisal rubric was designed and tested in advance of the search, to reflect the guiding questions being used in the review. Appraisal findings were carefully documented in an appraisal database. In the third, synthesis phase, the pool of highly relevant, high quality articles was methodically synthesized into themes and subthemes to develop an initial thematic structure for the review. This thematic structure was documented using thematic coding and in a set of structure diagrams and an initial thematic structure chart. Synthesis was followed by the fourth, analysis phase. In this phase the literature was analyzed in depth to create a comprehensive thematic map of the literature and to develop the insights needed to complete the literature review. The strength of the thematic connections between themes and subthemes was tested and refined, and additional germinal and connective articles along with professional texts were added to the thematic structure, as necessary to complete the analysis. Findings from this phase were updated into the thematic coding, article database, and thematic structure diagrams and charts, along with creation of a literature review database. This phase created a thematic structure to the literature and prepared the ground for the writing phase. In the final, writing phase, the outputs from earlier phases were carefully documented in this dissertation chapter. Figure 40 shows a context level flow chart of the SALSA process as used in this review.

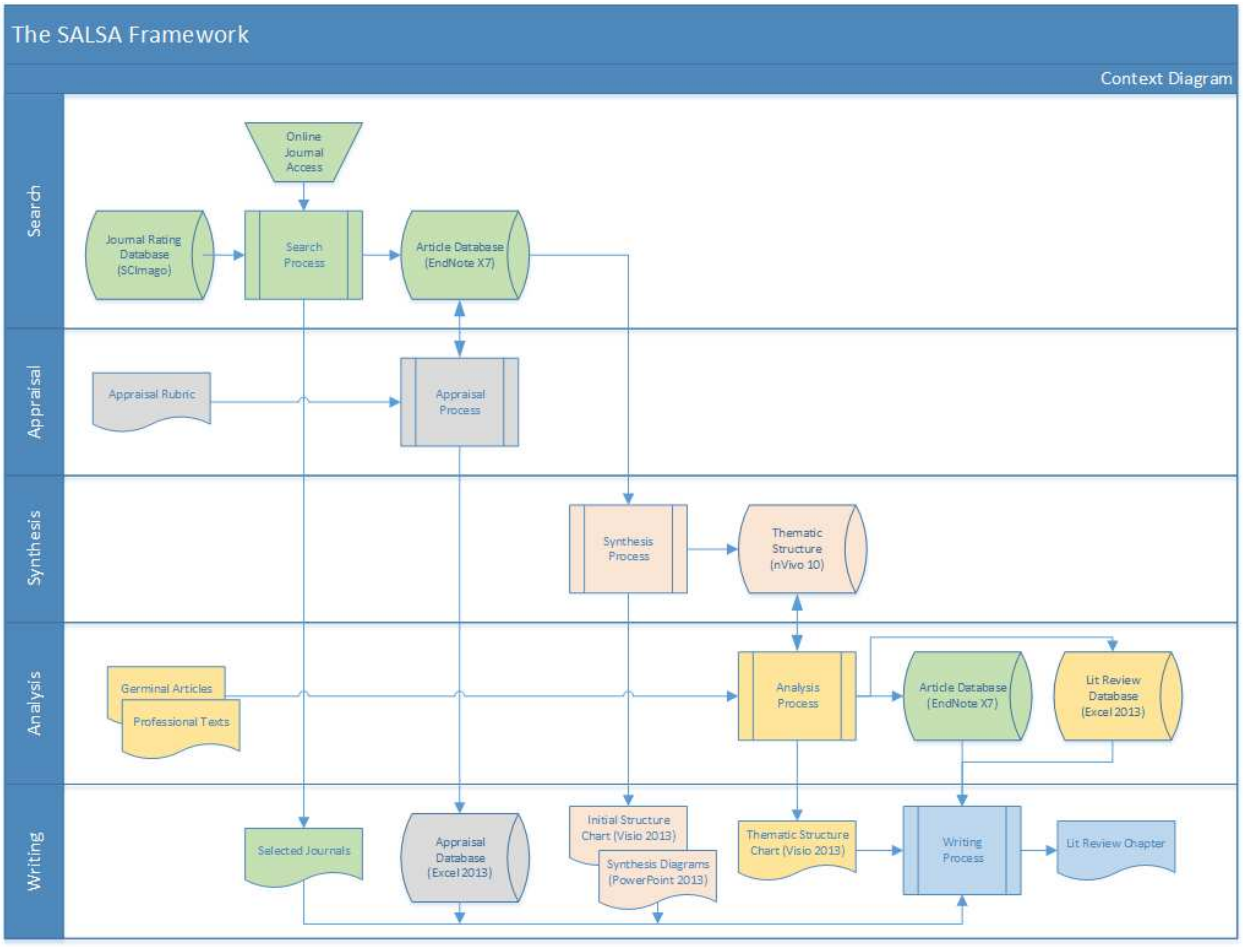


Figure 40. The SALSA Framework, adapted for use in this Review.

Notes. Adapted from Booth et al. (2012). Produced using Visio (Microsoft, 2013c) flowcharting software.

Journal Selection

In order to identify the most highly respected and relevant scholarly journals, I employed the Scimago Journal Rank (SJR2) indicator (Guerrero-Bote & Moya-Anegón, 2012). The SJR2 indicator uses a weighted measure to compute an index of relative journal prestige, based on citation counts per article, the closeness of the citations to the topics in question, and the proportion of citable (i.e. primary research) articles included in the journal. The SJR2 indicator is a significant improvement over other indicators of this type, and provides a stable and valid

relative measure of journal prestige. The most current analysis available is the 2012 SCImago Journal & Country Ranks, found on the internet at scimagojr.com (Guerrero-Bote & Moya-Anegón, 2012). Using the online search and ranking tools, I selected the most highly regarded, relevant scholarly journals in the academy for my systematic search in four categories:

1) General Education Research;

2) Educational Psychology Research;

3) Learning Environments Research; and,

4) Research into Creativity & Creative Problem Solving.

I examined both the SJR2 indicator and the number of citations by other scholarly journals, as part of the evaluation and selection of these journals. I then evaluated each candidate journal in turn, with the goal of ensuring that I had the top three ranked journals in each category.

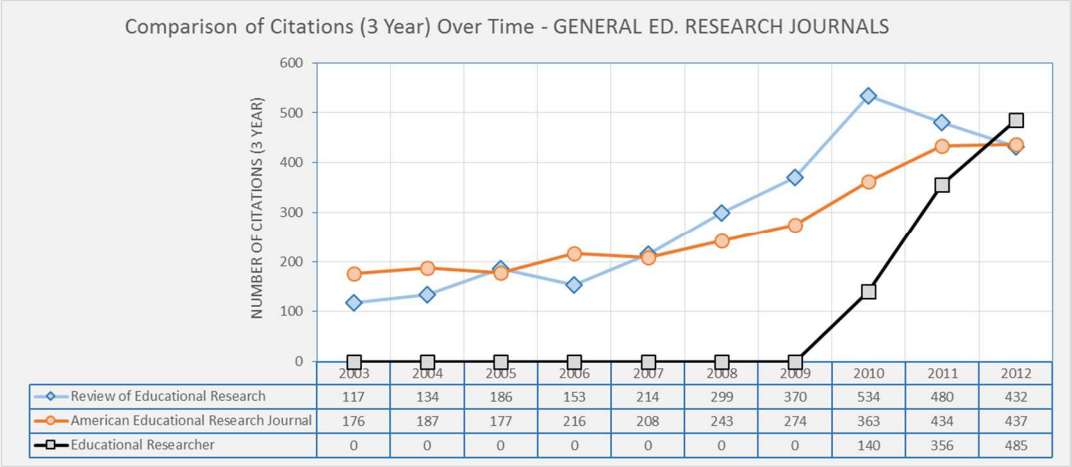
It became clear early in the evaluation of journals that the more general scholarly journals were more citable and had larger scholarly audiences due to their broader scope, and had therefore earned higher SJR2 scores. Because of this, the more narrowly defined journals (specifically journals focused on learning environments and creativity) did not yield three highly ranked scholarly journals, and scored significantly lower SJR2 scores overall. Rather than eliminating these categories of journals, I hypothesized that high quality research articles related to my guiding questions were likely to be published in the narrow journals as well as in the more general journals. I therefore decided to include all the candidate narrowly focused journals along with the top three journals in each general category, yielding a list of nine identified scholarly journals to be used in this review. Table 34 details the most prestigious scholarly journals (by SJR2 indicator) in each category.

Table 34.
Relevant Scholarly Journals, Ranked by SJR2 Indicator.

Category	Rank	Journal	SJR2[a]
General Educational Research	1	Review of Educational Research	3.86
	2	American Educational Research Journal	3.23
	3	Educational Researcher	2.52
Educational Psychology Research	1	Journal of Educational Psychology	2.81
	2	Contemporary Educational Psychology	2.51
	3	Journal of School Psychology	1.81
Learning Environments Research	1	Learning Environments Research	0.53
	2	n/a	-
	3	n/a	-
Research into Creativity & Creative Problem Solving	1	Creativity Research Journal	0.79
	2	Thinking Skills & Creativity	0.44
	3	n/a	-

[a]SJR2 indicator value for 2012.

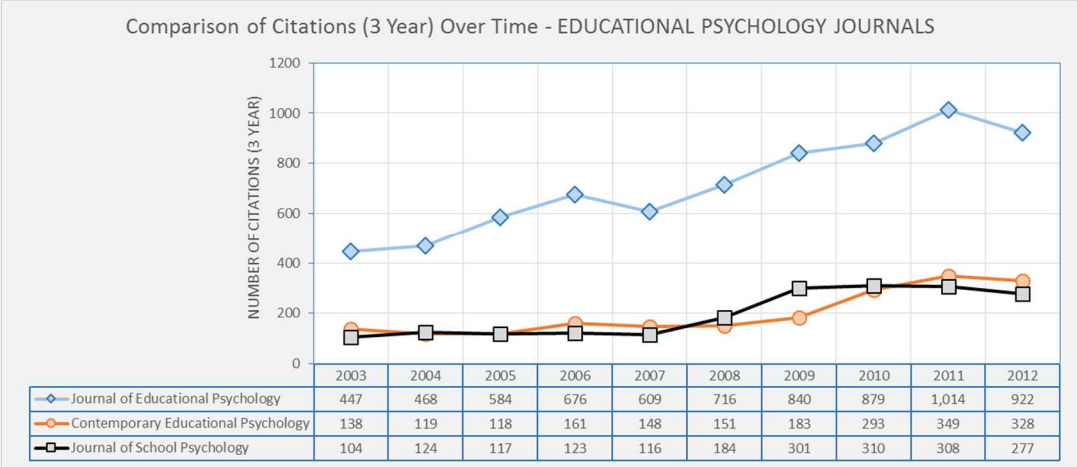
Analysis of the number of citations over a three year period confirmed the selection of these nine journals as the most highly regarded within their fields. Figures 41-44 show comparisons of citation counts for the selected journals.



	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Review of Educational Research	117	134	186	153	214	299	370	534	480	432
American Educational Research Journal	176	187	177	216	208	243	274	363	434	437
Educational Researcher	0	0	0	0	0	0	0	140	356	485

Figure 41. Prestigious General Educational Research Journals.

Note. Author's analysis of data extracted from (SCImago, 2007). Educational Researcher was not published between 2003 and 2008.



	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Journal of Educational Psychology	447	468	584	676	609	716	840	879	1,014	922
Contemporary Educational Psychology	138	119	118	161	148	151	183	293	349	328
Journal of School Psychology	104	124	117	123	116	184	301	310	308	277

Figure 42. Prestigious Educational Psychology Research Journals.

Note. Author's analysis of data extracted from (SCImago, 2007).

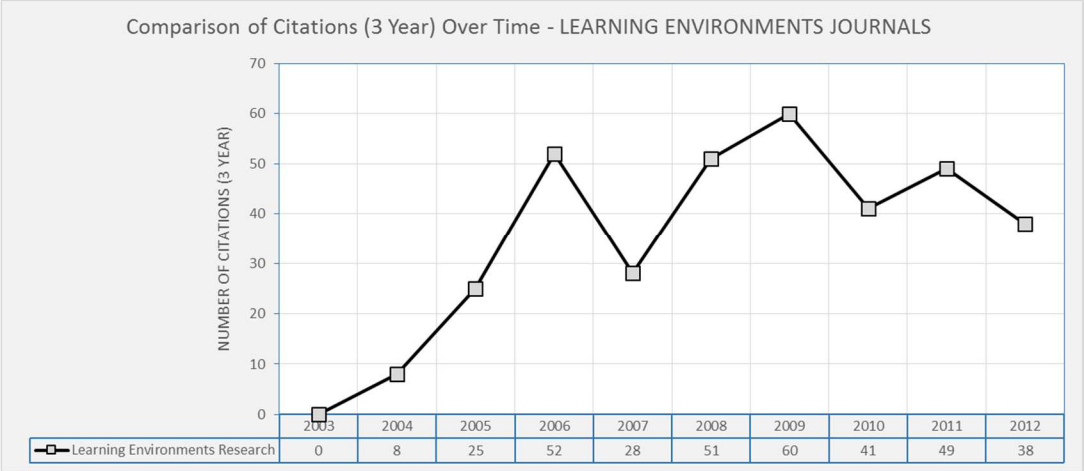


Figure 43. Prestigious Learning Environments Research Journals.

Note. Author's analysis of data extracted from (SCImago, 2007). Learning Environments Research was first published in 2003.

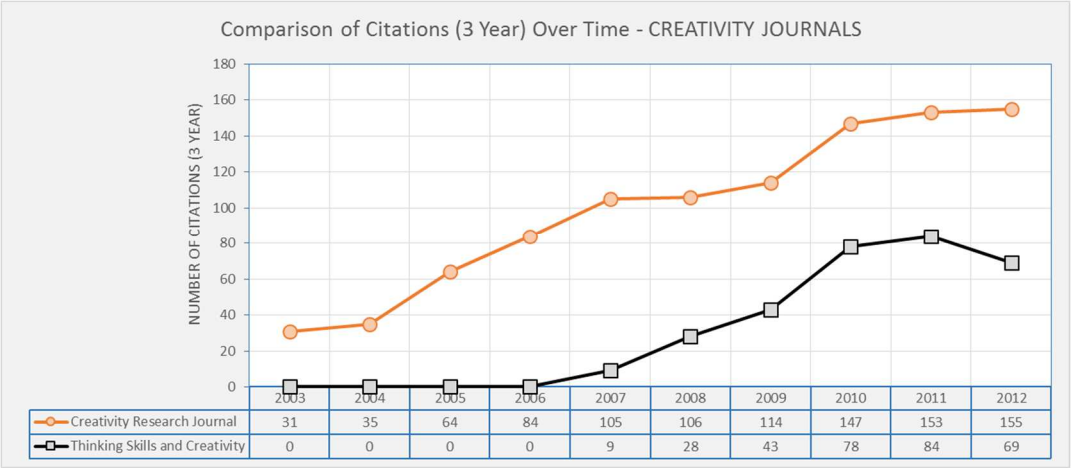


Figure 44. Prestigious Creativity Research Journals.

Note. Author's analysis of data extracted from (SCImago, 2007). Thinking Skills and Creativity was first published in 2006.

Search for Candidate Articles

Having identified the nine most highly regarded scholarly journals relevant to this review, I began my article search by iteratively trialing keyword searches across the universe of research articles found in these journals, specifically related to the guiding literature review

questions. These iterative trials led to the creation of two separate search streams: Search stream #1 focused on school climate and learning environments and search stream #2 focused on teaching and learning creativity & creative problem solving.

A total of 765 full text candidate articles were identified and downloaded from online databases or requested via interlibrary loan. Note that, particularly in conducting searches of Educational Researcher a large number (30) of non-research articles (member news and commentary &c.) were intermingled with research articles, despite the use of the "scholarly article" search criteria in my searches. Non-research articles found at this stage were excluded from the candidate article pool prior to becoming candidate articles. No other analysis of articles was conducted at this stage – all identified articles were treated as equal candidates for inclusion in the review.

Full text PDFs of all candidate articles were stored in an electronic folder located on my laptop computer and then loaded into *EndNote X7* software (ThomsonReuters, 2013). Secure real time backups of all materials were maintained using *Dropbox* software. Tables 35-36 show the search keywords used along with a count of candidate articles identified, by search stream and journal.

Table 35.
Search stream #1: School Climate & Learning Environments.

Keywords	Journal	Candidate article count
School Climate School Safety Bullying Anti-bullying Positive Behavior Social Emotional Learning PBIS SEL Learning Environment Classroom Environment School Community School Culture	Review of Educational Research	79
	American Educational Research Journal	154
	Educational Researcher	97
	Journal of Educational Psychology	48
	Contemporary Educational Psychology	117
	Journal of School Psychology	141
	Learning Environments Research	64
Total Candidate Articles Identified by Search Stream #1		**700**

Table 36.
Search Stream #2: Learning Creative Problem Solving.

Keywords	Journal	Candidate article count
Learning Creativity Learning to be Creative Learning Creative Problem Solving	Creativity Research Journal	21
	Thinking Skills & Creativity	44
Total Candidate Articles Identified by Search Stream #2		**65**

Appraisal of Candidate Articles

Having identified 765 candidate articles, I conducted a systematic critical appraisal, evaluating the relevancy and quality of each article in turn. This appraisal was guided by an objective appraisal rubric, shown in Table 37.

First, I examined the relevancy of each candidate article, using three criteria:

1. *Topical Relevancy* (i.e. was the article focused on topics within school climate, learning environments, or creative problem solving);
2. *Population Relevancy* (i.e. were participants American students in grades 3-8); and,
3. *Outcome Relevancy* (i.e. were researchers examining outcomes directly or indirectly relevant to my review questions).

This evaluation yielded three scores, each on a High (2) – Medium (1) – Low (0) scale (see Table 37 for narrative descriptions of each score), and a summative *Relevancy* appraisal scale of Highly Relevant (6-5) – Somewhat Relevant (4-3-2) – Not Relevant (1-0).

Next, each *highly relevant* article was assessed for quality, using four criteria:

1. *Paradigm* (i.e. did the article describe mixed or single method research);
2. *Research Design* (i.e. did the article clearly describe a high quality research design with strong fit to the stated research questions);
3. *Validation Arguments* (i.e. did the article contain a transparent, comprehensive, and paradigmatically appropriate argument for the validity of its findings); and,
4. *Actionability of Findings* (i.e. did the article contain specific, supportable, actionable recommendations for educational practice).

This evaluation yielded four scores, each on a High (2) – Medium (1) – Low (0) scale (see Table 37 for narrative descriptions of each score), and a summative *Quality* appraisal scale of High (8-7-6) – Medium (5-4-3) – Low (2-1-0).

The net effect of this critical appraisal approach was to deliberately privilege mixed methods research over single paradigm research, reflecting my own bias toward the use of multiple methods in programs of research. In addition this rubric served to treat explicit, well-formed validation arguments and actionable findings as essential structural components of a quality research study.

Table 37.
Critical Literature Appraisal Rubric.

Criteria	Sub criteria	High (2)	Medium (1)	Low (0)
Relevancy	Topical	Directly related to RQs.	Indirectly related to RQs.	Not related to RQs.
	Population	Study findings are for U.S. students between the ages of 8 and 11 (grades 3-13).	Study findings are for students between the ages of 4 and 18 (grades K-12).	Study does not include K-12 school students.
	Outcome	Examines one or more direct educational outcomes.	Examines one or more outcomes indirectly related to educational outcomes.	Study does not examine ed. outcomes.
Quality	Paradigm	Mixed methods design.	Single method design.	Not research-based.
	Research Design	High quality design, clear research questions.	Medium quality design, research questions included.	Poor quality design, no research questions.
	Validation Argument	High quality validity argument presented.	Partial validity argument presented.	Validity argument missing or poorly constructed.
	Action-ability	Findings imply specific recommended action or actions.	Findings imply new understanding of a phenomena.	Findings do not imply recommended action or understanding.

A total of 95 out of the 765 (12.4%) candidate articles were appraised as highly relevant, high quality articles. Table 38 summarizes the appraisal results for each journal included in the initial search. The *EndNote Rating* attribute was used to identify articles in the EndNote (ThomsonReuters, 2013) database, with the 5-star rating being used for highly relevant, high quality articles; 4-star rating for highly relevant, medium quality articles; 3-star rating being used for highly relevant, low quality articles; 2-star rating indicating medium relevancy articles; and a 1-star rating being used for low relevancy articles. The 0-star rating was used to indicate articles not yet rated. Summary scores were gathered at the sub-criteria level and entered into a Microsoft Excel (Microsoft, 2013a) appraisal database.

Table 38.
Summary Results from Appraisal Phase.

Journal	**Candidate article count**	**Highly relevant article count (%[1])**	**Highly relevant, high quality article count (%[1])**
Review of Educational Research	79	9 (11.4%)	7 (8.9%)
American Educational Research Journal	154	16 (10.3%)	15 (9.7%)
Educational Researcher	97	8 (8.2%)	6 (6.2%)
Journal of Educational Psychology	48	16 (33.4%)	8 (16.7%)
Contemporary Educational Psychology	117	16 (13.7%)	11 (9.4%)
Journal of School Psychology	141	43 (30.5%)	26 (18.4%)
Learning Environments Research	64	9 (14.1%)	8 (12.5%)
Creativity Research Journal	21	6 (28.5%)	4 (19.0%)
Thinking Skills & Creativity	44	12 (27.2%)	10 (22.7%)
Total	**765**	**135 (17.6%)**	**95 (12.4%)**

Note. [1]percentage of candidate articles.

Synthesis of Highly Relevant, High Quality Articles

Following the appraisal of candidate articles, the 95 highly relevant, high quality articles were loaded in *nVivo 10 for Windows* (QSR, 2012) software for synthesis. A series of initial autocode/review/synthesize cycles were completed, designed to progressively elaborate a cohesive thematic structure for this review. Each cycle developed a broad theme, along with a number of subthemes, and identified connection and separation points between the new theme and the existing thematic structure, thereby evolving the initial thematic structure of the literature.

In each autocode/review/synthesis cycle, word search queries were used to autocode articles into an emergent thematic structure. The word search queries themselves were developed by trial and error in advance of each cycle, using context free word trees to identify and refine the most productive word searches associated with each potential theme and subtheme. The most productive word trees were then converted into word searches and autocoded at appropriate nodes. Once this initial autocoded theme structure had been created, each node was reviewed to eliminate miscoding and to manually add additional sources. It was evident that the most significant uncoding required was removal of secondary sources, where these source names included one of the keywords used in my word searches. Since this is a review of primary research literature, sources which only reference keywords in citations in their references section were uncoded at these nodes.

Articles were then synthesized across subthemes within the theme, using matrix coding queries, with the initial 95 articles as rows, and the subthemes as columns. Connections were coded as necessary and a Venn diagram of the subthemes within each theme was created, allowing connections between subthemes to be explored. The nVivo cell color coding using the Blue-White option, allowed me to quickly identify where each source was connected to multiple

subthemes, and where a subtheme had multiple sources and high numbers of references within a particular source - identifying the most significant authors within each subtheme, as well as which subthemes were discussed in the literature as being related to one another. If an article included eight or more references to a particular subtheme, I treated that as a significant emphasis on a subtheme. Less than eight references were treated as a peripheral reference, and no references meant that an article did not include this subtheme.

This visual tool allowed me to refine the coding structure within each theme, and then enabled me to expand my synthesis across themes, as each new theme was added. On the second and subsequent themes, matrix coding queries were created showing the subthemes from two themes at a time, enabling a similar synthesis of cross-theme and cross-subtheme connections. Cross-theme Venn diagrams were produced, to help in this analysis. Figure 45 illustrates this process, showing the color coded Blue-White matrix developed following the creation of the first two major themes: safety and community.

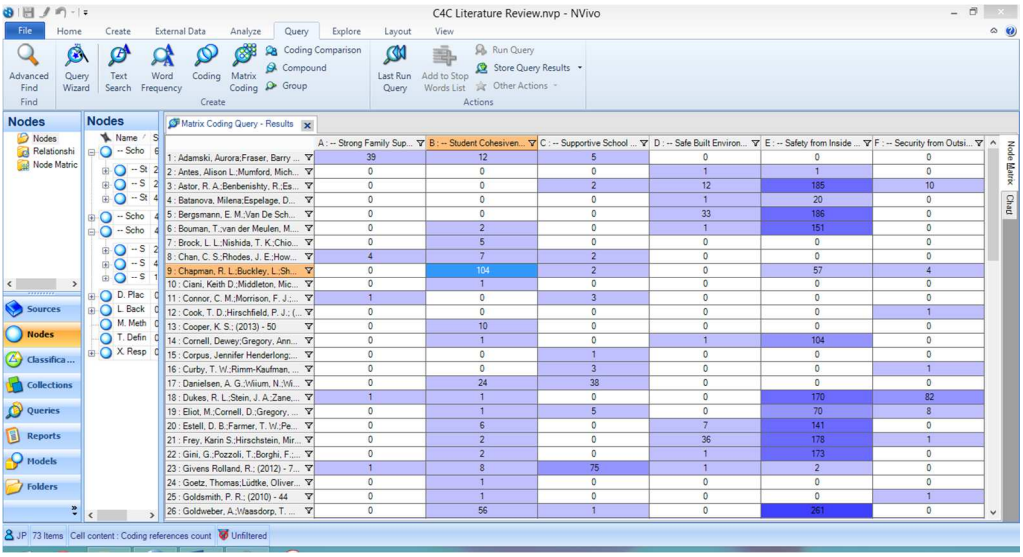


Figure 45. Color Coded Matrix Coding Query Results during Synthesis Phase.

Notes. Screen shot from nVivo 10 for Windows (QSR, 2012). The darker cells show the highest reference count, articles listed down the left, subthemes across the top. This allows the researcher to quickly identify connection points inside each theme and subthemes, and then extend the synthesis across themes and subthemes.

Safety. The first identified theme in this literature was physical and emotional safety, encompassing three subthemes:

1. Safety from inside threats (including research into bullying, aggression, and violence perpetuated by and acting on students or school staff);
2. Security from outside threats (including research into school shootings); and,
3. Safe built environment (including research into the physical buildings and the places and times when students felt more or less safe while at school).

Several articles focused within each subtheme, and a number of articles spanned subthemes. Figure 46 illustrates the initial connections identified between the three subthemes within the safety theme.

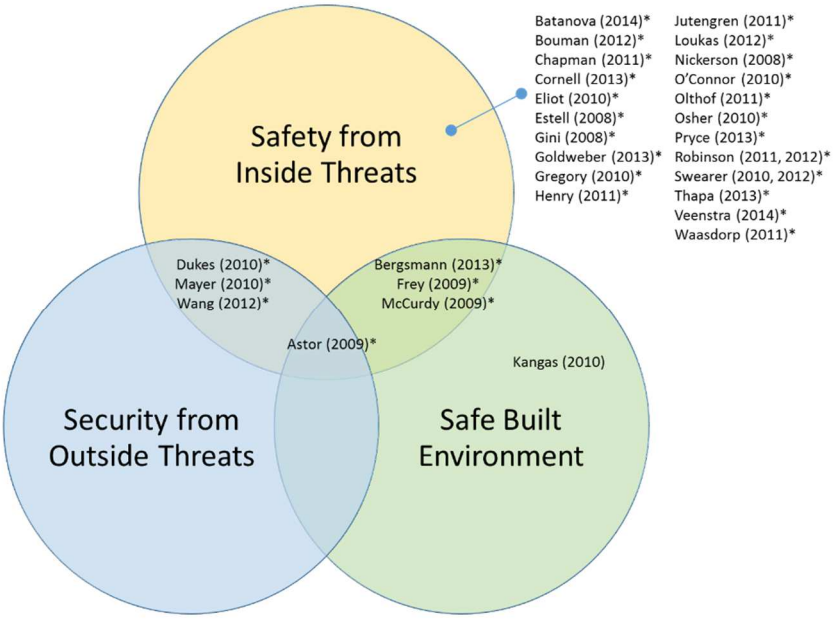


Figure 46. The Literature on Safety Organized[1] into Three Subthemes.

Note. * Articles with multiple authors identified by their first author. [1]initial thematic structure prior to analysis.

Since the majority of the quite extensive literature focused on the safety from inside threats subtheme, I further elaborated this subtheme into three subcomponents, providing a more granular organization of this part of the literature. Segments of the literature on safety from inside threats deal with implications and prevention of bullying, violence, and aggression and harassment in school contexts, as illustrated in Figure 47.

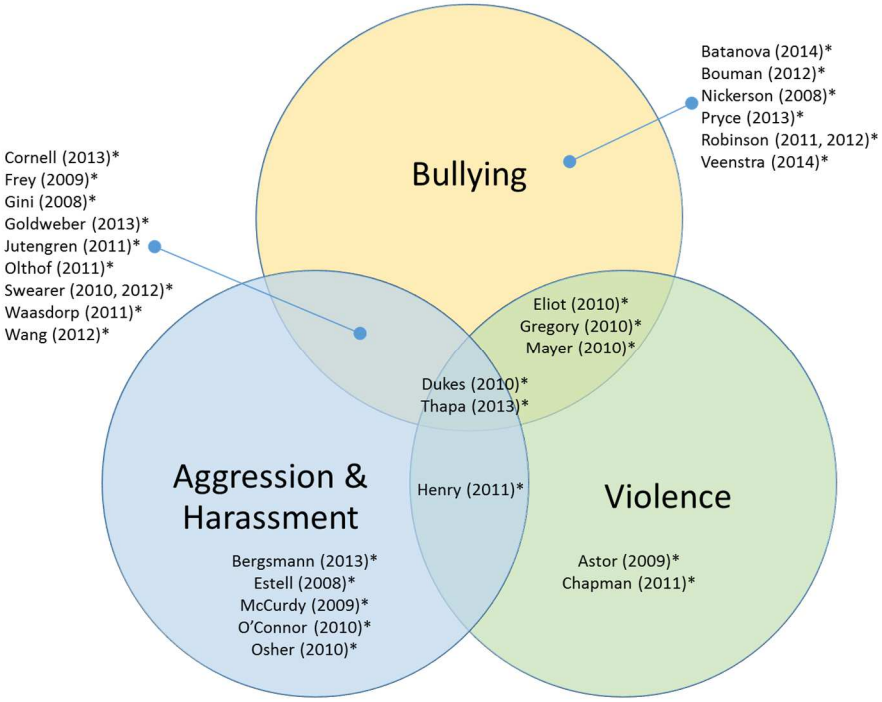


Figure 47. The Literature on Safety from Inside Threats Organized[1] into Three Subthemes.

Note. * Articles with multiple authors identified by their first author. [1]initial thematic structure prior to analysis.

Community. The second theme identified from the literature was school as a community, wrapped around the children studying at a school. This theme included three subthemes:

1. Family support (including school engagement, parenting, and family support for homework and other school activities);
2. Supportive school staff (including teacher-student relationships, and the degree of support provided by non-teaching staff at the school); and,
3. Student cohesiveness (particularly relatedness, social environment, and feelings of belonging in the school).

Once again, several articles focused within each subtheme, and a smaller number of articles spanned subthemes. Figure 48 illustrates the connections between the three subthemes identified within the community theme.

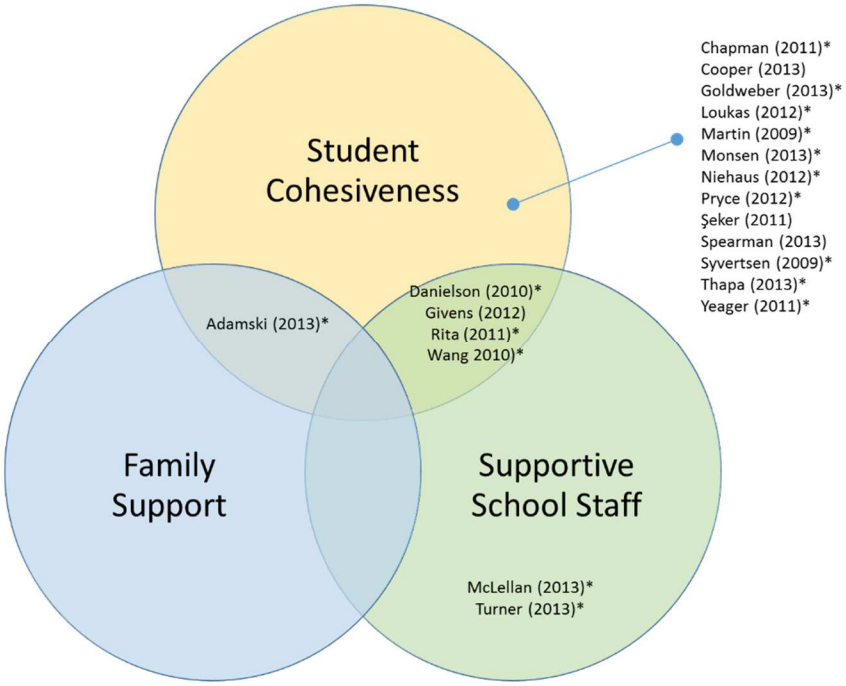


Figure 48. The Literature on Community Organized[1] into Three Subthemes.

Note. * Articles with multiple authors identified by their first author. [1]initial thematic structure prior to analysis.

Following synthesis of subthemes within the community theme, a cross-theme synthesis between community and safety was conducted, and two major connections were identified. First, Syvertsen et al. (2009) included significant references to both student cohesiveness and safety from outside threats, creating a potentially important cross-theme connection. Second, five research articles – Chapman et al. (2011), Goldweber et al. (2013), Gregory et al. (2010), Loukas et al. (2012), and Pryce and Frederickson (2013) – and a literature review of school climate (Thapa et al., 2013) connected student cohesiveness with safety from inside threats. Figure 49 illustrates the initial connections identified between the safety and community themes.

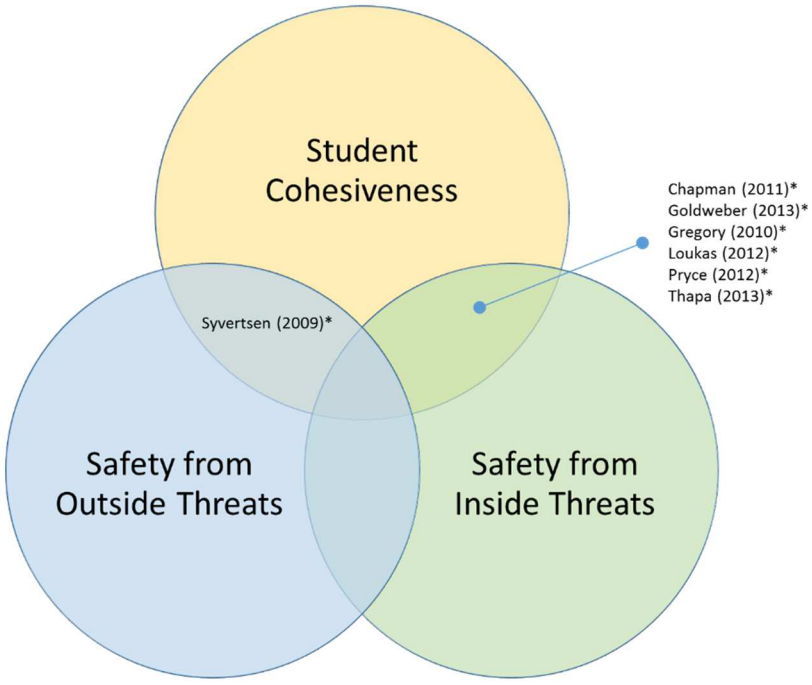


Figure 49. Connections[1] between the Literature on Safety and Community.

Note. * Articles with multiple authors identified by their first author. [1]initial thematic structure prior to analysis.

Learning Environments. The third theme found in the literature related to the environment for learning in a school. I coded this theme into four subthemes, as follows:

1. Diverse learning styles (including teaching methods, content, and environments);
2. High expectations (of self and perceived from teacher and family expectations);
3. Effective teaching (including time management, clear direction, organization); and,
4. Challenging curriculum.

Once again, the literature provided some material within each subtheme, and some articles that spanned one or more themes. In this case, it was interesting to note that the challenging curriculum subtheme appeared unrelated to the other three subthemes, in the literature. This warrants further analysis in the next phase. Figure 50 illustrates the connections between the four subthemes identified within the learning theme.

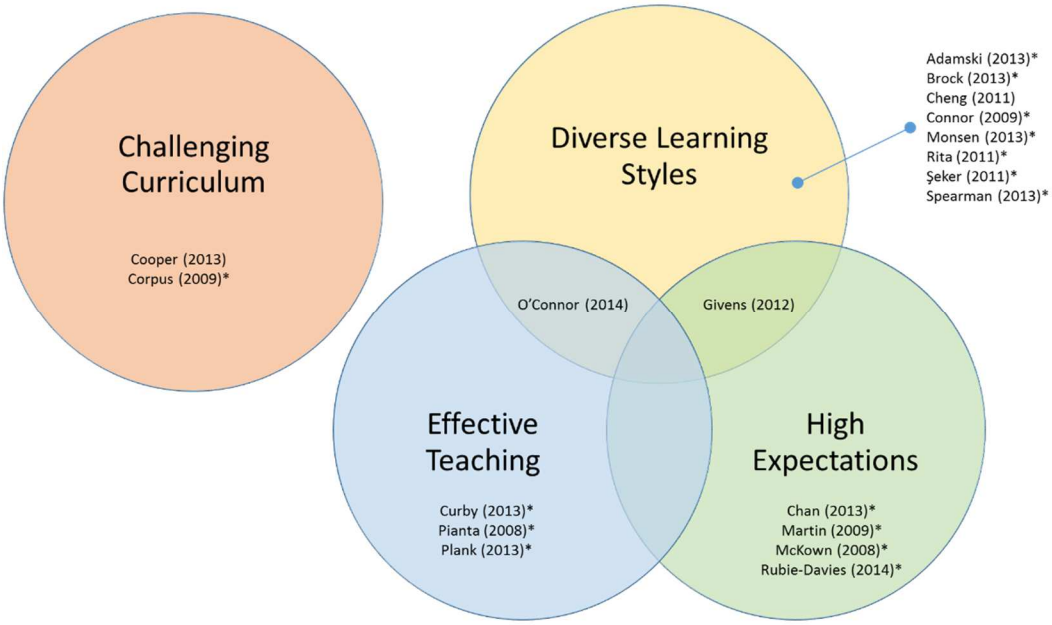


Figure 50. The Literature on Learning Environments Organized[1] into Four Subthemes.

Note. * Articles with multiple authors identified by their first author. [1]initial thematic structure prior to analysis.

Following synthesis of the literature within the learning theme, a cross-theme synthesis was conducted between the learning and safety themes, identifying no significant connections between these themes. A second cross-theme synthesis between learning and community, however, identified connections between six of the seven subthemes, including one significant connection between the challenging curriculum subtheme (which was not connected to the remaining learning subthemes) and the student cohesiveness subtheme, implying that challenging curriculum is more closely related to school community than it is to learning within a school. Figure 51 shows the connections between the literature within the community and learning themes.

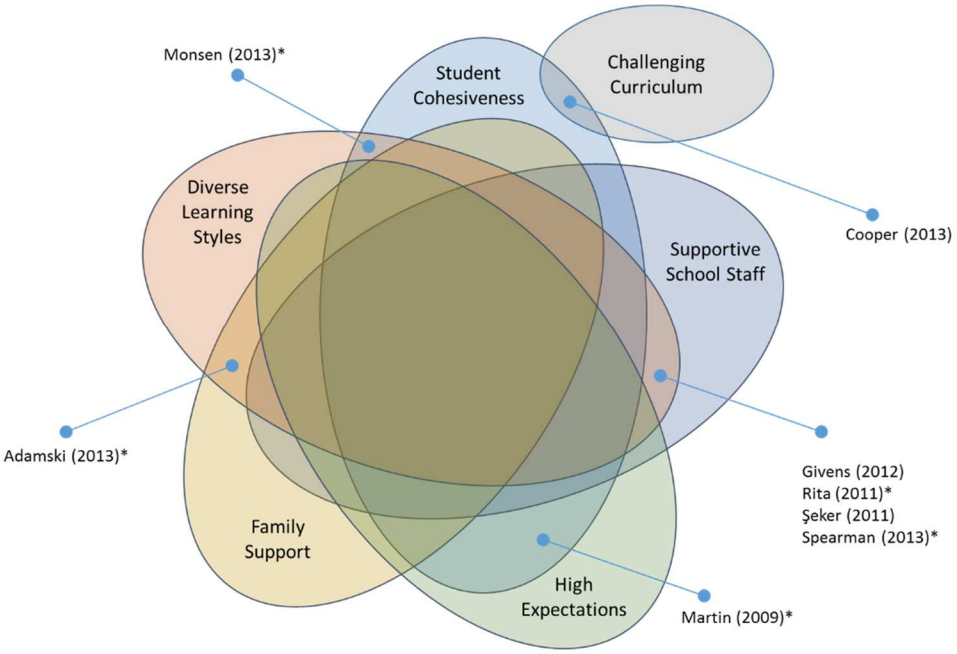


Figure 51. Connections[1] between the Literature on Learning Environments and Community.

Notes. * Articles with multiple authors identified by their first author. [1]initial thematic structure prior to analysis.

Creativity. The fourth theme identified in the literature refers to the teaching and learning of creativity and creative problem solving. This theme was elaborated into four subthemes:

1. Creative focus and flow (including enablers and barriers to focus and flow);
2. Freedom to create (teachers allowing students to be creative in school assignments);
3. Creative challenge (appreciation that creative problem solving is best employed in solving complex and broad challenges); and,
4. Support for creativity (both explicit via classroom goals and expectations, and implicit via recognition of creative work).

The literature provided some material within each subtheme, and some articles that spanned one or more themes. In this case, it was interesting that most of the literature was integrative in nature, combining identified subthemes in interesting ways. This warrants further analysis in the

next phase, as it was a noticeable difference versus the prior three themes. Figure 52, illustrates the connections between the subthemes identified within the creativity theme.

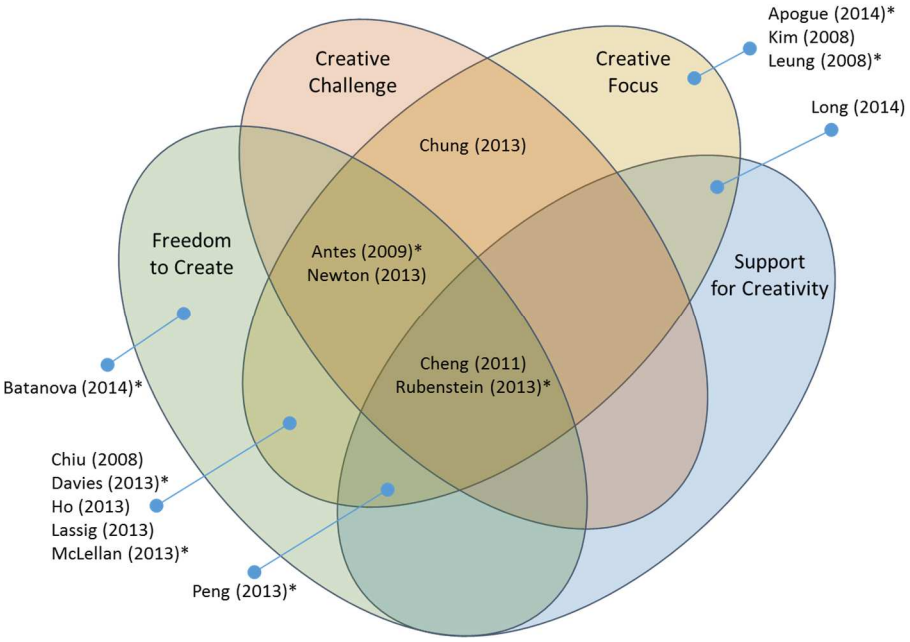


Figure 52. The Literature on Creativity Organized[1] into Four Subthemes.

Notes. * Articles with multiple authors identified by their first author. [1]initial thematic structure prior to analysis.

Following the subtheme synthesis within the creativity theme, a cross-theme synthesis was conducted between the creativity and learning themes, identifying a single connection between diverse learning styles and all four subthemes within creativity (Cheng, 2011). No other significant connections were initially identified between the creativity literature and learning literature. The cross-theme synthesis between the creativity and community themes found no significant direct connections, and the cross-theme synthesis between the creativity and safety theme found a single connection between the freedom to create subtheme and the safety from inside threats subtheme (Batanova et al., 2014).

Exhaustivity Review

Once all cycles were complete, the literature was reviewed for exhaustivity, to identify articles which had not been identified as making a significant reference to any of the subthemes identified in the synthesis cycles. This exhaustivity review mechanism was designed to identify any missing subthemes and themes in the thematic structure. First, I coded every source with eight or more keyword references at the subtheme level as *In Synthesis*, on the premise that articles with eight or more references to a subtheme were likely focused on topics related to that subtheme. This mechanism allowed me to identify twenty-three articles which had not been identified as significantly referencing any of the subthemes (i.e. which were not now coded as in synthesis).

These twenty three articles were individually reviewed and twenty were easily integrated into the existing subthemes (see Table 39). Two methodologically oriented articles, Lee (2010) and Lüdtke, Robitzsch, Trautwein, and Kunter (2009), were eliminated from the synthesis. One article (Goldsmith, 2010) referred to elements of race and culture and their impact on learning, leading to the addition of a new subtheme of respect for diversity, organized under the community theme.

Table 39.
Adjustments to Initial Thematic Structure based on Exhaustivity Review.

Theme	Articles Added to Theme	
Safety	Hand (2009)	
Community	Ciani et al. (2010) Lawson and Alameda-Lawson (2011) Mitchell and Bradshaw (2013) Peter and Dalbert (2010) Rudasill et al. (2010)	Cook and Hirschfield (2008) Mainhard et al. (2011) Patall et al. (2008) Roorda et al. (2011)
Learning Environment	Koth et al. (2008) Reis et al. (2010) Rimm-Kaufman et al. (2014)	Martin (2008) Reznitskaya et al. (2012) Wirkala and Kuhn (2011)
Teaching & Learning Creativity	Abrami et al. (2008) Mathews and Lowe (2011)	Goetz et al. (2013)

The iterative synthesis approach used in this phase of this literature review, along with the careful exhaustivity review described above, yielded a robust initial thematic structure to the literature, with four major themes, and a fifteen subthemes. The nVivo (QSR, 2012) toolkit was invaluable in completing this synthesis. Areas of the literature with conflicting findings were not excluded from the coding structure, as I viewed these as potential sources of insight later in the review. Figure 53, shows the initial synthesis map produced using the Visio software package (Microsoft, 2013c). The size of the bubbles indicates the depth of the literature within each subtheme, and the width of the connections indicates the strength of the connections between the subthemes.

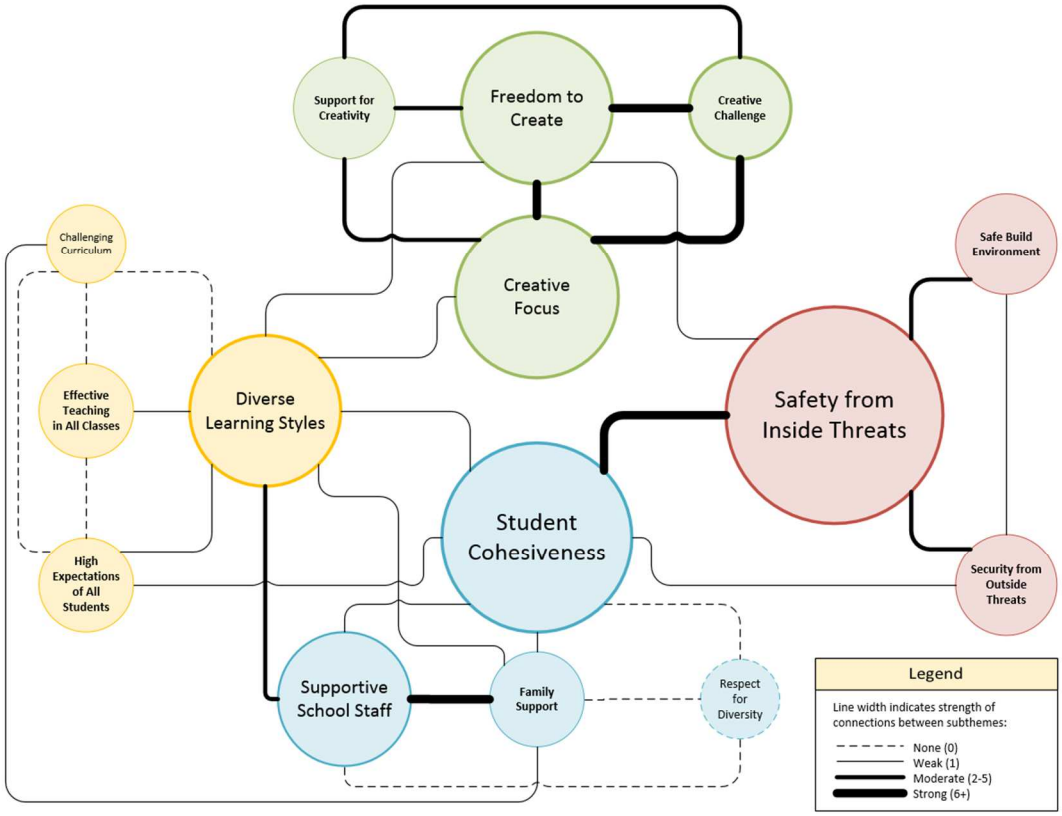


Figure 53. Initial Thematic Structure[1] Map of the Literature.

Notes. Size of bubbles and width of lines indicate strength of literature. [1]initial thematic structure prior to analysis.

Analysis of Articles and Themes

The penultimate phase in the SALSA process is the analysis phase. In this case, the 93 remaining articles were reviewed in detail, using the initial thematic structure as a framework and sequencing mechanism. The relative strength of each article was analyzed within each subtheme, and across subthemes, and updates to the initial thematic structure were made as I reviewed each article. The article strengths were entered into the literature review database as High, Medium, Low, or Zero for each subtheme, and any additional subthemes were identified and added to the thematic structure as needed. In order to connect subthemes and themes during

this process, nineteen connective scholarly articles and thirteen professional texts were identified from common citations within the initial literature pool. A full list of additional articles and texts added in this phase is provided in Table 40. Articles or texts which connected multiple subthemes were identified as germinal articles. Since these were selected as the important articles and texts for this review, they were included even if they were not as recent as the research articles included in the appraisal phase. Each article or text was individually appraised using the same objective appraisal rubric (Table 37), and if qualified, was synthesized into the thematic structure using manual coding.

Note that several of these additional sources were not available as full text electronic files, and so were dummy coded in order to maintain the integrity of the thematic structure. As each article was analyzed, brief narrative annotations were prepared in an electronic document. These narrative annotations provided the information required for the writing phase to follow this phase. The completed thematic structure diagrams and charts are presented in chapter 2 of this dissertation.

Table 40.
Additional Articles and Texts added during the Analysis Phase.

Theme	Connective Articles	Professional Texts
Safety	Astor and Meyer (2001) Astor et al. (2001) Borum et al. (2010) Hoover et al. (1993) Janssen et al. (2004) Juvonen et al. (2003) Little (2002) Marini et al. (2001) Nansel et al. (2003) Olweus (1995) Woods and Wolke (2004)	Eyman and Cohen (2009) Jimerson, Nickerson, Mayer, and Furlong (2012) O'Toole (2000) Vossekuil et al. (2002)
Community	Norwich and Kelly (2004)	National School Climate National School Climate Council (2007)
Learning Environment	Beer and Nohria (2000) Colvin et al. (1997)	Anderson and Krathwohl (2001) Marzano et al. (2003) Wallas (1926)
Teaching & Learning Creativity	Amabile et al. (2002) Beghetto and Kaufman (2009) DiLiello et al. (2011) Kaufman and Beghetto (2009) Mainemelis (2002) Renzulli (1992) Wallach and Kogan (1965)	Csikszentmihalyi (1990, 1996) Isaksen and Isaksen (2010) Kaufman and Sternberg (2010) Runco (2007) Sawyer (2006)

Writing the Literature Review

Once the four preparation phases were complete, the writing phase began with documentation of the methodology used in this review. Since every article had previously been loaded into EndNote (ThomsonReuters, 2013), along with both the appraisal and lit review

databases, creation of my bibliography and appendices was trivial. Next, the thematic structure and narrative annotations developed in the analysis phase were used to create the main findings section, and finally, the discussion section was prepared, designed to connect the findings back to the initial research questions and prepare the reader for entry into the primary research work of this dissertation.

Appendix C – An Informal Overview of Measurement Theory

When we measure something, we assign numerical values to observations. The simplest example, measurement of the length of an object using a ruler, illustrates this well. We place the ruler beside the object, match the scale on the ruler to the observed length of the object, and read the numerical value from the ruler. The ruler is an established and widely accepted *measurement instrument* for the length of an object. When a measurement instrument is first created, however, it is appropriate to exercise a degree of skepticism about the readings from the instrument, until its validity and reliability have been established through repeated use and analysis of its findings.

A *measure* of something is the numerical value on a scale read from a measurement instrument, which provides us with a way to develop observations about an underlying idea or *construct*, and thereby draw conclusions or make inferences about that underlying construct. In the physical sciences, we use *measurement instruments* (such as rulers and thermometers) to measure a particular property (such as length or temperature) of a physical object. In the human sciences, we use *psychometric instruments* (such as surveys and tests) to measure a particular property (such as attitudes or knowledge) of a human respondent. Unlike the use of instruments in the physical sciences, where we can often directly observe the usefulness of the measures by looking at the underlying property being measured, the use of psychometric instruments in the human sciences is, to use a technical term, *tricky*. We can never directly observe attitudes or knowledge – they are often quite abstract – we can, however, observe human behavior and find ways to measure that behavior as a proxy for the underlying attitudes or knowledge.

For example, in an elementary school, we might have students take a math test to assess their understanding of *addition and subtraction* – an abstract idea with a simple manifestation known as *computation*. In a rigorous math test, we would measure students' ability to compute a

series of selected problems and to explain their understanding of these problems by writing about the strategies they used to solve them. These computational and explicatory responses are *behavioral manifestations* of the underlying knowledge we aim to understand, and we feel comfortable using such measures to infer a measure of the students' true knowledge: in this case, if a student computes the answers correctly and appropriately explains her problem solving strategies we would conclude that she understands addition and subtraction. In reality, we go further, assuming the test score to actually *be* the student's understanding of the knowledge we aim to measure: she *is* an A, B, or C student in math; she *is* above, at, or below standards. We routinely treat this type of test measure as a *visible manifestation* of the *latent knowledge* we are attempting to understand. Wright and Stone (1979), in *Best Test Design*, describe this treatment,

> When we test a person, our purpose is to estimate their location on a line implied by the test. Before we can do this we must construct a test that defines a line. We must also have a way to turn the person's test performance into a location on that line. (Wright & Stone, 1979, p1)

Clearly, when we set out to measure something, we want to use a measurement instrument of an appropriate quality for the importance of the thing we intend to measure. We know what we are trying to do or learn with the measure, and would seek the appropriate instrument for our intended purpose. If we decide to develop a new instrument, we construct the instrument based on our understanding of our intended purpose, and as part of an integrated development and testing process we evaluate the quality of our new instrument in light of that intended purpose.

Rasch Measurement

The fundamental measurement problem in the human sciences is the quest for meaningful, rational, additive units of measure, units of measure which operate as effectively as

a ruler or thermometer operates in the physical sciences. We seek instruments in particular, with additive qualities, so that the distance between values on the instrument can be equated and compared in legitimate ways. No-one in her right mind would argue that the distance between, "Strongly Agree" and "Agree" is demonstrably equal to the distance between "Disagree" and "Neither Agree nor Disagree," and yet we routinely assign the values 5 (SA), 4 (A), 3 (N), 2 (D), 1 (SD) to the Likert scale and compute averages and distributions using these numerical assignments. Social scientists focus their considerable brain power on complex statistical analysis using these numerical assignments but pay little attention to the instruments used to collect categorical data in the first place, or to the mechanisms used to convert categories into numbers (Bond & Fox, 2007). Rasch measurement sets out to accomplish the task of assigning meaningful, interval measures to categories in a way that allows human scientists to extract meaning from an instrument at comparable levels to measurement in the physical sciences.

Historically, in the physical sciences, measures were seen as either fundamental (direct) measures such as length or mass or derived (indirect) measures such as density or temperature. In this paradigm, the derived measures need not be additive, since they are formed by the combination of two or more fundamental measures, which are themselves additive. However, if we were to combine two linear, interval measures, into a matrix, showing the derived measure at the intersection of row and column values, we could examine the relationships between the values in the middle of the matrix, and begin to assign distances between the derived values, based on the combination of the fundamental values on the axes of the matrix (Bond & Fox, 2007).

In 1960, Georg Rasch (Rasch, 1960) published such a matrix for students taking tests. Rasch showed test *items* across the top of his matrix, ordered by *difficulty* of the items, and

persons taking the test down the side of his matrix, ordered by their *ability*. He argued that a person of average ability should have a 50/50 chance (0.500 probability) of getting a problem of average difficulty correct, and that all the other probabilities were predictable, too. This item-person matrix remains at the heart of Rasch measurement today – and the family of measurement models derived from Rasch's work are the, "closest generally accessible approximation of these fundamental measurement principles for the human sciences." (Bond & Fox, 2007, p. 14)

So, how does this all work? On an educational test, we would create a matrix of items and persons, modeled after Rasch's matrix. Each cell (combination of test item and test person) would either be a correct answer or an incorrect answer (a 1 or a 0). Table 41 provides an illustrative example of this type of matrix.

Table 41.
Illustrative example of a Test Data Matrix for Eight Persons and Seven Items.

	Items							
Persons	a	b	c	d	e	f	g	Raw Score
A	✓	✓	✓	×	×	×	×	3
B	✓	×	✓	✓	×	×	×	3
C	✓	✓	✓	×	✓	×	×	4
D	✓	×	✓	✓	✓	✓	×	5
E	×	✓	✓	×	×	×	×	2
F	×	×	×	×	×	×	×	0
G	✓	×	✓	×	×	×	×	2
H	✓	✓	✓	✓	✓	✓	✓	7

We could add up the total correct answers for each row (person) and for each column (item), and sort our matrix so that the highest ability person (i.e. the person who scored the most correct answers) was at the top of the matrix, and the easiest problem (i.e. the item the most people got right) was on the left. This format is known as a *scalogram*, and is illustrated in Table 42.

Table 42.
Illustrative example of a Scalogram based on the Test Data Matrix.

	Items							
Persons	c	a	b	d	e	f	g	Ability
H	✓	✓	✓	✓	✓	✓	✓	7
D	✓	✓	×	✓	✓	✓	×	5
C	✓	✓	✓	×	✓	×	×	4
A	✓	✓	✓	×	×	×	×	3
B	✓	✓	×	✓	×	×	×	3
E	✓	×	✓	×	×	×	×	2
G	✓	✓	×	×	×	×	×	2
F	×	×	×	×	×	×	×	0
Facility	7	6	4	3	3	2	1	

Just the act of sorting test data into a scalogram begins to reveal some useful information about the test and about the persons taking the test. In the case of a test, we want a range of items from easy to hard, so that we can meaningfully discriminate between the ability of persons taking the test. Note that this does not necessarily mean having some items that nobody can answer and some items than everybody can answer. In the example, we cannot accurately make judgments about the ability of person H, other than to say that the test was too easy for her. Likewise, we cannot make an accurate judgment about the ability of person F, other than to say the test was too hard for him. For now, we would remove these persons from our analysis, since we cannot place them on the scale. Next time we test our population, we should make sure there are a wider array of problem difficulties, so that we can locate an estimate for the true ability of all our test takers.

The next step is to convert the raw ability and facility scores into percentages of the total possible scores on each axis. We can now begin to qualitatively examine *patterns* in the matrix, and draw some early conclusions about what is going on with both items and persons. In

development of a new instrument, we seek anomalies and qualities in the items in order to provide meaningful measures of the persons taking the test. One issue with the use of raw scores is that, while we may be able to effectively differentiate between persons in the middle of the range of ability, and between items in the middle of the range of difficulty, it may be difficult to understand the differences at the low and high ends of the ability and facility scales. To resolve this issue, a simple mathematical computation called a *log-odds* calculation can be performed to convert raw scores expressed as probabilities into their *logit* equivalent. This has the net impact of stretching the raw scale, and reducing floor and ceiling effects. This mathematical technique is comparable to techniques using in the physical sciences when calculating derived measures.

The underlying logit computation used in Rasch measurement is a very simple calculation. It begins with the odds ratio – the probability of a person achieving their raw score divided by the probability of that same person not achieving their raw score. So, a score of 45% becomes an odds ratio of 45%/55% = 45/55 = ~0.82. The logit score is simply the natural log of the odds ratio, as follows:

$$\text{logit}\ (\ p\ ) = \log_e \left\{ {}^{p}\!/_{(1\ -\ p)} \right\}$$

Where *p* is the raw score expressed as a probability.

Applying this computation to the raw scores from our illustrative example (ignoring persons H and F), gives us a logit scale as shown in Table 43. Note that the logit scale is centered at 0, and, on a larger test, typically ranges from around -4.0 to +4.0.

Table 43.
Logit Scale based on Raw Person Ability Scores.

Person	Raw Score	Person Ability (p)	Odds Ratio = p / (1-p)	Logit (p)
D	5	5/7 = 71%	71/29	+0.90
C	4	4/7 = 57%	57/43	+0.28
A	3	3/7 = 43%	43/57	-0.28
B	3	3/7 = 43%	43/57	-0.28
E	2	2/7 = 29%	29/71	-0.90
G	2	2/7 = 29%	29/71	-0.90

Rasch Measurement in Instrument Development

Once a theory to be examined with a measure has been converted into one or more constructs, a set of items are defined, designed to measure a line or *pathway*, corresponding to the measure of the theory. The instrument developer constructs items that fall along a continuum of difficulty, from easy to difficult, and that test as many aspects of the theory as practical. If the theory is based on the work on one individual, the items may be constructed based on a chapter from a single text written by that individual. In the case of the theory of Climate4Creativity, the items were constructed based on a review of the literature on school safety, school community, general learning environments, and the environment for teaching and learning creativity.

As items are developed they are first designed to be as individually effective as possible by understanding the respondents' cognitive processes, related to the subject matter of the construct being measured. In this case, a cognitive lab was used to refine and improve items by interviewing a representative sample of respondents with similar demographics to the target population for the measurement instrument. Wording and answer choices are refined and adapted to ensure that each item is individually clear and meaningful to the target respondent population. Each item in the measure should contribute in some meaningful way to the researcher's understanding of the respondent's feelings about the construct being measured.

This concept is known as *unidimensionality*, and essentially sets an expectation that each item measures one thing only, and measures it in the same direction. Items should be connected to one and only one construct, and item scales should be oriented in the same direction as the construct scale (Bond & Fox, 2007).

Once a reasonable set of items have been defined, they are pretested with another sample of the population, and a *pathway analysis* is completed, to examine the *fit* of the items to the Rasch model overall. Pathway analysis extends the illustration shown in Table 43, to include elements of the distribution of person responses at each level of the item scale. This allows the instrument developer to understand and improve the quality of the items design by finding items which do not fit the model, because their mean response and distribution falls outside of the zone (or pathway) determined by the model. Figure 54 shows an illustrative pathway analysis. Note that a logit scale is shown at the center of the pathway diagram, and the locations of items on the scale is shown on the pathway chart, based on a similar logic to the person-ability calculations shown above. The width of the white section in the center of the diagram (the pathway) is, $t = \pm 2s$, and represents an acceptable amount of variability of item-person responses to items in the measure. Items which fall inside the pathway are considered to fit the model, and those which fall outside, such as item h in the example, do not fit the model. Item fit is viewed as one of the major quality control checks for a measurement scale.

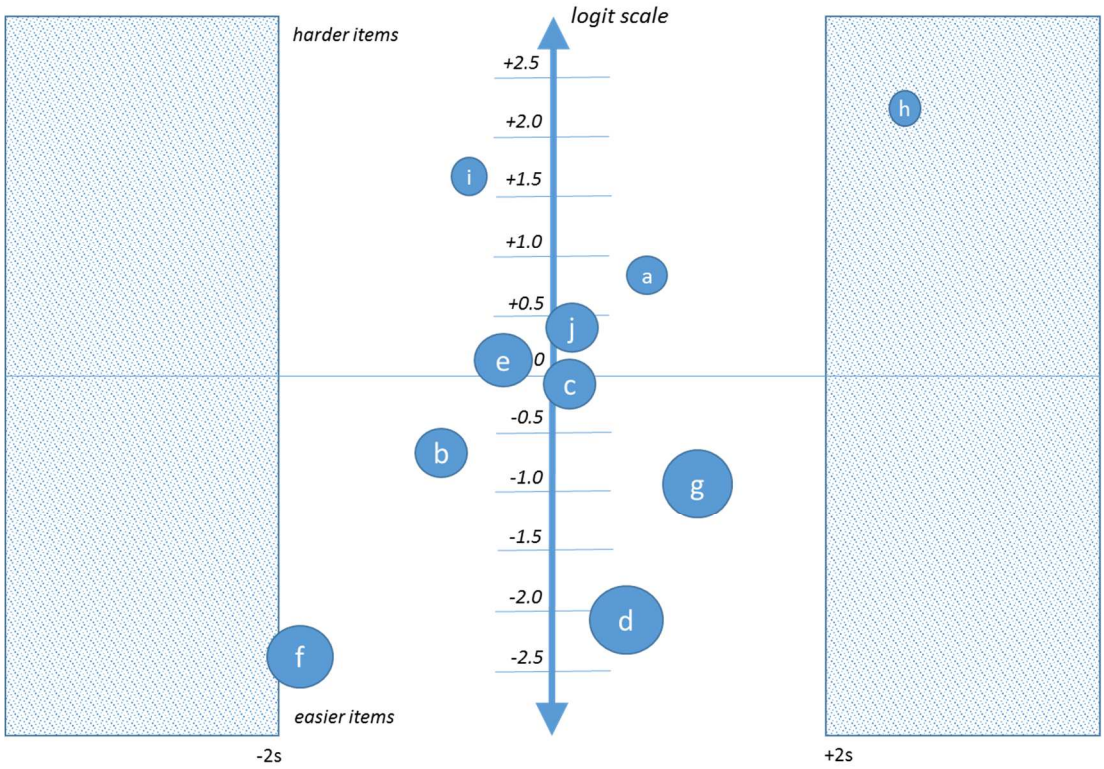


Figure 54. Illustrative Pathway for a Measurement Scale.
Adapted from Bond and Fox (2007), cover illustration.

In Rasch analysis, we examine the pretest data, comparing it to Rasch measurement model expectations, to determine if the actual data fits the expectations of the Rasch model. If the data fit, we can begin to conclude things about the actual levels of respondents in the measure. If the data do not fit, we cannot conclude much from the data. In other words, the Rasch measurement model, applied to pretest pilot data, can be used as a form of empirical evidence of the validity of the underlying constructs of the measure, as well as a way of empirically establishing the quality of the items design itself. This is fundamental to how measurement instruments are constructed using Rasch measurement models (Bond & Fox, 2007).

Appendix D – The Climate4Creativity v1.1 (Pilot Study) Items Design

The Climate4Creativity Student Perspectives Instrument, Elementary v1.1

Table 44.
Climate4Creativity Student Perspectives Instrument, Elementary v1.1

Item	Item Text	Answer Choices (Numerical Values)	Construct
1-GR	What grade are you in?	Grade 3 (3) Grade 4 (4) Grade 5 (5) Grade 6 (6)	Demographics
2-GNDR	Are you a girl or a boy?	Girl (1) Boy (2)	Demographics
3-RACE	What is your race?	White (1) Black/African American (2) Native American (3) Asian American (4) Pacific Islander (5) Other Race (6)	Demographics
4-HISP	Are you Hispanic/Latino?	Yes – I am Hispanic / Latino (1) No – I am NOT Hispanic / Latino (2)	Demographics
5-ENGL	Do you speak English at home?	Yes – I speak English at home (1) No – I don't speak English at home (2)	Demographics
How are your family members involved in school			
6A-HWK	They help me with my homework	Always (4) Usually (3) Sometimes (2) Never (1)	Community
6B-RPT	They read my report cards		
6C-WKHD	They tell me to work hard		
6D-TTCH	They talk with my teachers		
6E-EVNT	They come to school events		
How much do people help you to learn at school?			
7A-FRND	My closest school friends help me	Always (4) Usually (3) Sometimes (2) Never (1)	Community
7B-CLS	Other students in my class help me		
7C-STU	Students in other grades help me		
7D-TCHR	My teachers help me		
7E-ADLT	Other adults in my school help me		
How do people treat each other at your school?			
8A-KNSS	Students are kind to other students	Always (4) Usually (3) Sometimes (2) Never (1)	Community
8B-RSSS	Students are respectful to other students		
8C-KNTS	Teachers are kind to students		
8D-RSTS	Teachers are respectful to students		
8E-KNST	Students are kind to teachers		
8F-RSST	Students are respectful to teachers		
9-LIKE	How much do you like your school?	Love My School (4) Like My School (3) My School is OK (2) Hate My School (1)	Community

Cont/…

Item	Item Text	Answer Choices (Numerical Values)	Construct
How safe do you usually feel during your school day?			
10A-TRAV	Traveling to and from school	Very Safe (4) Safe (3) A Little Unsafe (2) Very Unsafe (1)	Safety
10B-B4AF	At school before and after class		
10C-CLS	In class with my teacher there		
10D-RCSS	At recess or at lunchtime		
How safe do you usually feel in different places at school?			
11A-CLS	In my classroom	Very Safe (4) Safe (3) A Little Unsafe (2) Very Unsafe (1)	Safety
11B-HALL	Walking in the halls		
11C-BATH	Going to the bathroom		
11D-PLGR	Outside in the playground		
11E-STRS	Walking up or down the stairs		
11F-PLOT	Outside in the parking lot		
11G-LRM	In the lunch room		
How does your school stop people coming in?			
12A-LOCK	The outside doors are locked	Always (4) Usually (3) Sometimes (2) Never (1)	Safety
12B-STKR	Visitors have to wear stickers		
12C-BUZZ	People have to buzz in		
12D-CHK	The security guards check people		
How well does your school stop bullying?			
13A-HURT	They keep me safe from getting hurt	Always (4) Usually (3) Sometimes (2) Never (1)	Safety
13B-STFF	They keep my stuff safe		
13C-AFRD	They help me not be afraid at school		
13D-NICE	They make school a nice place		
How much bullying do you see or hear about at your school?			
14A-TALK	People talk about bullying (ʁ)	Every Day (1) Most Days (2) Sometimes (3) Never (4)	Safety
14B-SEE	I see people bullying others (ʁ)		
15-BULL	Have you ever been bullied at school? (ʁ)	I am being bullied now (1) I was before but it has stopped now (2) I have never been bullied (3) There is no bullying at my school (4)	Control
What do adults do when students tell them about bullying?			
16A-STOP	They stop the bullying	Always (4) Usually (3) Sometimes (2) Never (1)	Safety
16B-HELP	They help the person who is hurt		
16C-PUN	They punish the bully		
16D-TALK	They talk to the class about bullying		

Cont/…

Item	Item Text	Answer Choices (Numerical Values)	Construct
What do students do when they see bullying?			
17A-HELP	They help the person who is hurt	Always (4) Usually (3) Sometimes (2) Never (1)	Safety
17B-STOP	They tell the bully to stop		
17C-TELL	They tell an adult what happened		
18-SAFE	How safe do you feel at school overall?	Very Safe (4) Safe (3) A Little Unsafe (2) Very Unsafe (1)	Safety
19-EXP	What does your teacher expect from you?	Great Work (4) OK Work (3) Bad Work (2) Nothing (1)	Learning
Think about your teacher…			
20A-DIR	My teacher gives clear directions	Always (4) Usually (3) Sometimes (2) Never (1)	Learning
20B-ENCR	Encourages me to be creative		Creativity
20C-ENHL	Encourages me to help others		Learning
20D-RECR	Recognizes me for being creative		Creativity
20E-TIME	Gives me time to finish my work		Learning
Think about the work you do in class…			
21A-FUN	The work I do in class is fun	Always (4) Usually (3) Sometimes (2) Never (1)	Creativity
21B-THNK	Makes me think about stuff		
21C-DIFF	I get to work on different things		Learning
21D-HARD	The work I do is difficult		
21E-TEAM	I work on things with other students		
How much do you…			
22-LRN	Learn new things in class	Always (4) Usually (3) Sometimes (2) Never (1)	Learning
23-CRE	Be creative in class		Creativity

Note. (ʁ) Reversed scale.

The Climate4Creativity Student Perspectives Instrument, Middle School v1.1

Table 45.
Climate4Creativity Student Perspectives Instrument, Middle School v1.1

Item	Item Text	Answer Choices (Numerical Values)	Construct
1-GR	What grade are you in?	Grade 6 (6) Grade 7 (7) Grade 8 (8)	Demographics
2-GNDR	What is your gender?	Female (1) Male (2)	Demographics
3-RACE	What is your racial identity?	White (1) Black/African American (2) Native American (3) Asian American (4) Pacific Islander (5) Other Race (please tell us) (6)	Demographics
4-HISP	Are you Hispanic/Latino?	Yes – I am Hispanic / Latino (1) No – I am NOT Hispanic / Latino (2)	Demographics
5-ENGL	Is English you first language?	Yes – English is my first language (1) No – English is NOT my first language (2)	Demographics
6-FAVE	What is your favorite subject at school?	[school specific list]	Control
7-LFAV	What is your LEAST favorite subject at school?	[school specific list]	Control
How much are members of your family involved in your school?			
8A-HWK	They help me with my homework	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Community
8B-RPT	They read my report cards		
8C-WKHD	They encourage me to work hard		
8D-TTCH	They talk with my teachers about me		
8E-EVNT	They come to school events		
8F-PTO	They are involved in a parents' organization		
How much do different people help, support and encourage you to learn at school?			
9A-FRND	My closest school friends help me	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Community
9B-CLS	Other students in my class help me		
9C-STU	Students in other grades help me		
9D-TCHR	My teachers help me		
9E-ADLT	Other adults in my school help me		
How do students treat each other at school?			
10A-KNSS	Students are kind to other students	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Community
10B-RSSS	Students are respectful to other students		
10C-KNTS	Teachers are kind to students		
10D-RSTS	Teachers are respectful to students		
10E-KNST	Students are kind to teachers		
10F-RSST	Students are respectful to teachers		

Cont/…

Item	Item Text	Answer Choices (Numerical Values)	Construct
11-LIKE	Overall, how much do you like your school?	Love My School (5) Like My School (4) School is OK (3) Dislike My School (2) Hate My School (1)	Community
How safe do you usually feel during your school day?			
12A-TRAV	Traveling to school	Completely Safe (5) Very Safe (4) Mostly Safe (3) Somewhat Unsafe (2) Very Unsafe (1)	Safety
12B-B4	Before classes start for the day		
12C-CLS	In class with my teacher there		
12D-LNCH	At lunchtime		
12E-AFTR	After classes end for the day		
12F-HOME	Traveling home		
How safe do you usually feel in different places at school?			
13A-CLS	In my classroom	Completely Safe (5) Very Safe (4) Mostly Safe (3) Somewhat Unsafe (2) Very Unsafe (1)	Safety
13B-HALL	Walking in the corridors		
13C-BATH	Going to the bathroom		
13D-PLGR	Outside in the playground		
13E-STRS	Walking up or down the stairs		
13F-PLOT	Outside in the parking lot		
13G-LRM	In the lunch room		
How well does your school stop people coming in?			
14A-LOCK	The outside doors are locked	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Safety
14B-STKR	Visitors have to wear stickers		
14C-BUZZ	People have to buzz in		
14D-CHK	The security guards check people		
How well does your school stop bullying?			
15A-HURT	They keep me safe from getting hurt	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Safety
15B-STFF	They keep my stuff safe		
15C-AFRD	They help me not be afraid at school		
15D-NICE	They make school a nice place		
15E-TCH	They teach us about bullying		
How much bullying happens at your school?			
16A-TALK	People talk about bullying (ʁ)	Every Day (1) Most Days (2) Sometimes (3) Not Often (4) Never (5)	Safety
16B-SEE	I see people bullying others (ʁ)		
17-BULL	Have you ever been bullied at school? (ʁ)	I am being bullied now (1) [+] I was before but it has stopped now (2) [+] I have never been bullied (3) There is no bullying at my school (4)	Safety, Control

Cont/…

Item	Item Text	Answer Choices (Numerical Values)	Construct
What do adults do when you tell them about bullying?			
18A-STOP	They stop the bullying	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Safety
18B-HELP	Help the person who is hurt or upset		
18C-PUN	They punish the bully		
18D-TALK	They talk to the class about bullying		
What do students do when they see bullying?			
19A-HELP	They help the person who is hurt	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Safety
19B-STOP	They tell the bully to stop		
19C-TELL	They tell an adult what happened		
20-SAFE	How safe do you feel at your school overall?	Completely Safe (5) Very Safe (4) Mostly Safe (3) Somewhat Unsafe (2) Very Unsafe (1)	Safety
21-FGRX	What grade do you usually earn in [F] class?	A (5) B (4) C (3) D (2) F (1)	Learning
22A-FEXP	What does your [F] teacher expect from you?	Excellent Work (5) Good Work (4) OK Work (3) Poor Work (2) Expects Me To Fail (1)	Learning
22B-FHRD	How difficult is the work you do in [F] class?	Very Difficult (5) Somewhat Difficult (4) About Right (3) Somewhat Easy (2) Very Easy (1)	Learning
Thinking about your [F] teacher...			
23A-FDIR	My teacher gives clear directions	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Learning
23B-FNCR	Encourages me to be creative		Creativity
23C-FRTH	Lets me rethink or reframe class problems		
23D-FHLP	Encourages me to help others		Learning
23E-FREC	Recognizes me for being creative		Creativity
23F-FINV	Lets me invent my own ways to solve problems		
23G-FTIM	Gives me time to finish my work		Learning
23H-FINS	Inspires me to do better		

Cont/…

Item	Item Text	Answer Choices (Numerical Values)	Construct
Thinking about your [F] class...			
24A-FFUN	The work I do in class is fun	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Creativity
24B-FTHK	Makes me think about things		
24C-FDIF	I get to do different things		Learning
24D-FLTT	Makes me lose track of time		Creativity
24E-FTEM	Makes me work in teams with others		Learning
25-FCRE	How important do you think it is to be creative in [F] class? (ʁ)	We are not allowed to be creative in this class (1) Being creative in NOT a good thing (2) It is OK to be creative in this class (3) Being creative in this class is a good thing (4) It is very important to be creative in this class (5)	Creativity
26-FLRN	Overall, how much do you feel you are learning in [F] class? (ʁ)	I haven't learned anything in this class (1) I am not learning much in this class (2) I am learning most of what my teacher wants (3) I am learning exactly what my teacher wants (4) I am learning more than my teacher expects me to (5)	Learning
27-LGRX	What grade do you usually earn in [LF] class?	A (5) B (4) C (3) D (2) F (1)	Learning
28-LEXP	What does your [LF] teacher expect from you?	Excellent Work (5) Good Work (4) OK Work (3) Poor Work (2) Expects Me To Fail (1)	Learning
29-LHRD	How difficult is the work you do in [LF] class?	Very Difficult (5) Somewhat Difficult (4) About Right (3) Somewhat Easy (2) Very Easy (1)	Learning

Cont/…

Item	Item Text	Answer Choices (Numerical Values)	Construct
Thinking about your [LF] teacher...			
30A-LDIR	My teacher gives clear directions	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Learning
30B-LNCR	Encourages me to be creative		Creativity
30C-LRTH	Lets me rethink or reframe class problems		
30D-LHLP	Encourages me to help others		Learning
30E-LREC	Recognizes me for being creative		Creativity
30F-LINV	Lets me invent my own ways to solve problems		
30G-LTIM	Gives me time to finish my work		Learning
30H-LINS	Inspires me to do better		
Thinking about your [LF] class...			
31A-LFUN	The work I do in class is fun	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Creativity
31B-LTHK	Makes me think about things		
31C-LDIF	I get to do different things		Learning
31D-LLTT	Makes me lose track of time		Creativity
31E-LTEM	Makes me work in teams with others		Learning
32-LCRE	How important do you think it is to be creative in [LF] class? (ʁ)	We are not allowed to be creative in this class (1) Being creative in NOT a good thing (2) It is OK to be creative in this class (3) Being creative in this class is a good thing (4) It is very important to be creative in this class (5)	Creativity
33-LLRN	Overall, how much do you feel you are learning in [LF] class? (ʁ)	I haven't learned anything in this class (1) I am not learning much in this class (2) I am learning most of what my teacher wants (3) I am learning exactly what my teacher wants (4) I am learning more than my teacher expects me to (5)	Learning

Note. (ʁ) Reversed scale. [F] Selection option from 6-FAVE dropped into item text. [LF] Selected option from 7-LFAV dropped into item text. [+] Additional non-measure items based on selected option.

Additional Non-measurement Items, Middle School v1.1

Table 46.
Middle School v1.1. Additional Non-measurement Items.

Item Text	Answer Choices (Multi-select)
If 17-BULL = 1 (I am being bullied now):	
Tell us about your experience being bullied…	
Who is bullying you? [1]	Someone from outside the school A student from another grade level A student from another class in my grade level A student in my class One of my teachers Another adult in the school Someone else (please tell us who) [2]
Why do you think the bully is picking on you? [1]	Because I'm disabled Because of my racial/ethnic background Because of my gender Because they think I am gay Because I don't have the money or things they have Because of how I dress Because of my school work Some other reason (please tell us why) [2]
If 17-BULL = 2 (I was before but it has stopped now):	
Tell us about your experience being bullied in the past…	
Who bullied you? [1]	Someone from outside the school A student from another grade level A student from another class in my grade level A student in my class One of my teachers Another adult in the school Someone else (please tell us who) [2]
Why do you think the bully picked on you? [1]	Because I'm disabled Because of my racial/ethnic background Because of my gender Because they think I am gay Because I don't have the money or things they have Because of how I dress Because of my school work Some other reason (please tell us why) [2]

Notes. [1]Respondent may select multiple options. [2]Free form response is available.

Appendix E – The Climate4Creativity v1.2 (Calibration Study) Measure

The Climate4Creativity Elementary Student Instrument (C4C/SPE) v1.2e

Table 47.
Climate4Creativity Elementary Student Perspectives (C4C/SPE) v1.2e

Item	Item Text	Answer Choices (Numerical Values)	Construct
1-GR	What grade are you in?	Grade 3 (3) Grade 4 (4) Grade 5 (5)	Demographics
2-GNDR	Are you a girl or a boy?	Girl (1) Boy (2)	Demographics
3-HISP	Are you Hispanic/Latino?	Yes – I am Hispanic / Latino (1) No – I am NOT Hispanic / Latino (2)	Demographics
4-ENGL	Do you usually speak English at home?	Yes – We usually speak English at home (1) No – We do not usually speak English at home (2)	Demographics
5-RACE	What is your racial identity?[1]	White (1) Black/African American (2) Native American (American Indian) (3) Asian American (4) Pacific Islander (5) Other Race (6)[2]	Demographics
How much are your family members involved in your school?			
6A-HWK	They help me with my homework	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Community
6B-RPT	They read my report cards		
6C-WKHD	They encourage me to work hard		
6D-TTCH	They talk with my teachers about me		
6E-EVNT	They come to school events		
How much do different people help, support and encourage you to learn at school?			
7A-FRND	My closest school friends help me	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Community
7B-CLS	Other students in my class help me		
7C-STU	Students in other grades help me		
7D-TCHR	My teachers help me		
7E-ADLT	Other adults in my school help me		
How do people treat each other at your school?			
8A-KNSS	Students are kind to other students	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Community
8B-RSSS	Students are respectful to other students		
8C-KNTS	Teachers are kind to students		
8D-RSTS	Teachers are respectful to students		
8E-KNST	Students are kind to teachers		
8F-RSST	Students are respectful to teachers		
9-LIKE	Overall, how much do you like your school?	Love My School (5) Like My School (4) My School is OK (3) Dislike My School (2) Hate My School (1)	Community

Cont/…

Item	Item Text	Answer Choices (Numerical Values)	Construct
How safe do you usually feel during your school day?			
10A-TRAV	Traveling to and from school	Very Safe (5) Safe (4) Mostly Safe (3) Somewhat Unsafe (2) Very Unsafe (1)	Safety
10B-B4	Before classes start for the day		
10C-CLS	In class with my teacher there		
10D-LNCH	At lunchtime		
10E-RCSS	At recess		
10F-AFTR	After classes end for the day		
How safe do you usually feel in different places at school?			
11A-CLS	In my classroom	Very Safe (5) Safe (4) Mostly Safe (3) Somewhat Unsafe (2) Very Unsafe (1)	Safety
11B-HALL	Walking in the hallways		
11C-BATH	Going to the bathroom		
11D-PLGR	Outside in the playground		
11E-STRS	Walking up or down the stairs		
11F-PLOT	Outside in the parking lot		
11G-LRM	In the lunch room		
How does your school stop people coming in?			
12A-LOCK	The outside doors are locked	Very Safe (5) Safe (4) Mostly Safe (3) Somewhat Unsafe (2) Very Unsafe (1)	Control
12B-STKR	Visitors have to wear stickers		
12C-BUZZ	People have to buzz in		
12D-CHK	The security guards check people		
How well does your school stop bullying?			
13A-HURT	They keep me safe from getting hurt	Very Safe (5) Safe (4) Mostly Safe (3) Somewhat Unsafe (2) Very Unsafe (1)	Safety
13B-STFF	They keep my stuff safe		
13C-AFRD	They help me not be afraid at school		
13D-NICE	They make school a nice place		
13E-TCH	They teach us about bullying		
14-TALK	How often do you hear students talk about being mean to others? (R)	Every Day (1) Most Days (2) Sometimes (3) Not Often (4) Never (5)	Safety
15-SEE	How often do you see people bullying others at your school (R)		
16-BULL	Have you ever been bullied at school? (R)	I am being bullied now (1) I was before but it has stopped now (2) I have never been bullied (3) There is no bullying at my school (4)	Control

Cont/…

Item	Item Text	Answer Choices (Numerical Values)	Construct
What do adults do when students tell them about bullying?			
17A-STOP	They stop the bullying	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Safety
17B-HELP	They help the person who is hurt or upset		
17C-PUN	They punish the bully		
17D-TALK	They talk to the class about bullying		
What do students do when they see bullying?			
18A-HELP	They help the person who is hurt	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Safety
18B-STOP	They tell the bully to stop		
18C-TELL	They tell an adult what happened		
19-SAFE	How safe do you feel at your school overall?	Very Safe (5) Safe (4) Mostly Safe (3) Somewhat Unsafe (2) Very Unsafe (1)	Safety
20-EXP	What does your teacher expect from you?	Excellent Work (5) Good Work (3) OK Work (3) Poor Work (2) Expects Me To Fail (1)	Learning
21-DIFF	How difficult is the work you do in class?	Very Difficult (5) Somewhat Difficult (4) About Right (3) Somewhat Easy (2) Very Easy (1)	Detail
Think about your teacher…			
22A-DIR	My teacher gives clear directions	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Learning
22B-ENCR	Encourages me to be creative		Creativity
22C-RETH	Lets me rethink or reframe class problems		
22D-ENHL	Encourages me to help others		Learning
22E-RECR	Recognizes me for being creative		Creativity
22F-INV	Lets me invent my own ways to solve problems		
22G-TIME	Gives me time to finish my work		Learning
22H-INSP	Inspires me to do better		

Cont/…

Item	Item Text	Answer Choices (Numerical Values)	Construct
Think about the work you do in class…			
23A-FUN	The work I do in class is fun	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Creativity
23B-THNK	Makes me think about stuff		
23C-DIFF	I get to do different things		Learning
23D-TIME	Makes me lose track of time		Creativity
23E-ACT	Involves different kinds of activities		Learning
23F-TEAM	Makes me work in teams with others		
23G-CREA	Includes being creative		Creativity
24-CRE	How important do you think it is to be creative in class? (ʁ)	We are not allowed to be creative in class (1) Being creative is NOT a good thing in class (2) It is OK to be creative in class (3) Being creative in class is a good thing (4) It is very important to be creative in class (5)	Creativity
25-LRN	How often do you feel you learn new things in class?	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Learning

Note. (ʁ) Reversed scale. [1]Respondent may select multiple options. [2]Free form text box provided if this option checked.

The Climate4Creativity Middle School Student Instrument (C4C/SPM) v1.2m

Table 48.
Climate4Creativity Middle School Student Perspectives (C4C/SPM) v1.2m

Item	Item Text	Answer Choices (Numerical Values)	Construct
1-GR	What grade are you in?	Grade 6 (6) Grade 7 (7) Grade 8 (8)	Demographics
2-GNDR	Are you a girl or a boy?	Girl (1) Boy (2)	Demographics
3-HISP	Are you Hispanic/Latino?	Yes – I am Hispanic / Latino (1) No – I am NOT Hispanic / Latino (2)	Demographics
4-ENGL	Do you usually speak English at home?	Yes – We usually speak English at home (1) No – We do not usually speak English at home (2)	Demographics
5-RACE	What is your racial identity?[1]	White (1) Black/African American (2) Native American (American Indian) (3) Asian American (4) Pacific Islander (5) Other Race (please tell us) (6)[2]	Demographics
6-FAVE	What is your favorite subject at school?	[school specific list]	Control
7-LFAV	What is your LEAST favorite subject at school?	[school specific list]	Control
How much are your family members involved in your school?			
8A-HWK	They help me with my homework	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Community
8B-RPT	They read my report cards		
8C-WKHD	They encourage me to work hard		
8D-TTCH	They talk with my teachers about me		
8E-EVNT	They come to school events		
8F-PTO	They are involved in a parents' organiz'n		Detail
How much do different people help, support and encourage you to learn at school?			
9A-FRND	My closest school friends help me	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Community
9B-CLS	Other students in my class help me		
9C-STU	Students in other grades help me		
9D-TCHR	My teachers help me		
9E-ADLT	Other adults in my school help me		
How do people treat each other at your school?			
10A-KNSS	Students are kind to other students	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Community
10B-RSSS	Students are respectful to other students		
10C-KNTS	Teachers are kind to students		
10D-RSTS	Teachers are respectful to students		
10E-KNST	Students are kind to teachers		
10F-RSST	Students are respectful to teachers		

Cont/…

Item	Item Text	Answer Choices (Numerical Values)	Construct
11-LIKE	Overall, how much do you like your school?	Love My School (5) Like My School (4) My School is OK (3) Dislike My School (2) Hate My School (1)	Community
How safe do you usually feel during your school day?			
12A-TRAV	Traveling to and from school	Very Safe (5) Safe (4) Mostly Safe (3) Somewhat Unsafe (2) Very Unsafe (1)	Safety
12B-B4	Before classes start for the day		
12C-CLS	In class with my teacher there		
12D-LNCH	At lunchtime		
12F-AFTR	After classes end for the day		
How safe do you usually feel in different places at school?			
13A-CLS	In my classroom	Very Safe (5) Safe (4) Mostly Safe (3) Somewhat Unsafe (2) Very Unsafe (1)	Safety
13B-HALL	Walking in the hallways		
13C-BATH	Going to the bathroom		
13D-PLGR	Outside in the playground		
13E-STRS	Walking up or down the stairs		
13F-PLOT	Outside in the parking lot		
13G-LRM	In the lunch room		
How well does your school stop people coming in?			
14A-LOCK	The outside doors are locked	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Detail
14B-STKR	Visitors have to wear stickers		
14C-BUZZ	People have to buzz in		
14D-CHK	The security guards check people		
How well does your school stop bullying?			
15A-HURT	They keep me safe from getting hurt	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Safety
15B-STFF	They keep my stuff safe		
15C-AFRD	They help me not be afraid at school		
15D-NICE	They make school a nice place		
15E-TCH	They teach us about bullying		
16-TALK	How often do you hear students talk about being mean to others? (R)	Every Day (1) Most Days (2) Sometimes (3) Not Often (4) Never (5)	Safety
17-DRAM	How often do you see drama happening at your school? (R)		
18-SEE	How often do you see people bullying others at your school? (R)		

Cont/…

Item	Item Text	Answer Choices (Numerical Values)	Construct
19-BULL	Have you ever been bullied at school? (ʁ)	I am being bullied now (1) [+] I was before but it has stopped now (2) [+] I have never been bullied (3) There is no bullying at my school (4)	Control
What do adults do when students tell them about bullying?			
20A-STOP	They stop the bullying	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Safety
20B-HELP	They help the person who is hurt or upset		
20C-PUN	They punish the bully		
20D-TALK	They talk to the class about bullying		
What do students do when they see bullying?			
21A-HELP	They help the person who is hurt	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Safety
21B-STOP	They tell the bully to stop		
21C-TELL	They tell an adult what happened		
22-SAFE	How safe do you feel at your school overall?	Very Safe (5) Safe (4) Mostly Safe (3) Somewhat Unsafe (2) Very Unsafe (1)	Safety
23-FGRD	What grade do you usually earn in [F] class?	A (5) B (4) C (3) D (2) F (1)	Learning
24-FEXP	What does your [F] teacher expect from you?	Excellent Work (5) Good Work (3) OK Work (3) Poor Work (2) Expects Me To Fail (1)	Learning
25-FDIF	How difficult is the work you do in [F] class?	Very Difficult (5) Somewhat Difficult (4) About Right (3) Somewhat Easy (2) Very Easy (1)	Detail

Cont/…

Item	Item Text	Answer Choices (Numerical Values)	Construct
Think about your [F] teacher…			
26A-FDIR	My teacher gives clear directions	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Learning
26B-FENC	Encourages me to be creative		Creativity
26C-FRET	Lets me rethink or reframe class problems		
26D-FENH	Encourages me to help others		Learning
26E-FREC	Recognizes me for being creative		Creativity
26H-FINV	Lets me invent my own ways to solve problems		
26F-FTIM	Gives me time to finish my work		Learning
26G-FINS	Inspires me to do better		
Think about the work you do in [F] class…			
27A-FFUN	The work I do in class is fun	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Creativity
27B-FTHK	Makes me think about stuff		
27C-FDIF	I get to do different things		Learning
27D-FTIM	Makes me lose track of time		Creativity
27E-FACT	Involves different kinds of activities		Learning
27F-FTM	Makes me work in teams with others		
27G-FCRE	Includes being creative		Creativity
28-FCRE	How important do you think it is to be creative in [F] class? (ʁ)	We are not allowed to be creative in class (1) Being creative is NOT a good thing in class (2) It is OK to be creative in class (3) Being creative in class is a good thing (4) It is very important to be creative in class (5)	Creativity
29-FLRN	Overall, how much do you feel you are learning in [F] class? (ʁ)	I haven't learned anything in this class (1) I am not learning much in this class (2) I am learning most of what my teacher wants (3) I am learning exactly what my teacher wants (4) I am learning more than my teacher expects me to (5)	Learning
30-LGRD	What grade do you usually earn in [LF] class?	A (5) B (4) C (3) D (2) F (1)	Learning
31-LEXP	What does your [LF] teacher expect from you?	Excellent Work (5) Good Work (3) OK Work (3) Poor Work (2) Expects Me To Fail (1)	Learning

Cont/…

Item	Item Text	Answer Choices (Numerical Values)	Construct
32-LDIF	How difficult is the work you do in [LF] class?	Very Difficult (5) Somewhat Difficult (4) About Right (3) Somewhat Easy (2) Very Easy (1)	Detail
Think about your [LF] teacher…			
33A-LDIR	My teacher gives clear directions	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Learning
33B-LENC	Encourages me to be creative		Creativity
33C-LRET	Lets me rethink or reframe class problems		
33D-LENH	Encourages me to help others		Learning
33E-LREC	Recognizes me for being creative		Creativity
33H-LINV	Lets me invent my own ways to solve problems		
33F-LTIM	Gives me time to finish my work		Learning
33G-LINS	Inspires me to do better		
Think about the work you do in [LF] class…			
34A-LFUN	The work I do in class is fun	Always (5) Usually (4) Sometimes (3) Not Often (2) Never (1)	Creativity
34B-LTHK	Makes me think about stuff		
34C-LDIF	I get to do different things		Learning
34D-LTIM	Makes me lose track of time		Creativity
34E-LACT	Involves different kinds of activities		Learning
34F-LTM	Makes me work in teams with others		
34G-LCRE	Includes being creative		Creativity
35-LCRE	How important do you think it is to be creative in [LF] class? (ʁ)	We are not allowed to be creative in class (1) Being creative is NOT a good thing in class (2) It is OK to be creative in class (3) Being creative in class is a good thing (4) It is very important to be creative in class (5)	Creativity
36-LLRN	Overall, how much do you feel you are learning in [LF] class? (ʁ)	I haven't learned anything in this class (1) I am not learning much in this class (2) I am learning most of what my teacher wants (3) I am learning exactly what my teacher wants (4) I am learning more than my teacher expects me to (5)	Learning

Note. (ʁ) Reversed scale. [1]Respondent may select multiple options. [2]Free form text box provided if this option checked. [F] Favorite subject from question 6-FAVE dropped into question text. [LF] Least favorite subject from question 7-LFAV dropped into question text. [+] Additional non-measure items based on selected option.

Additional Non-measurement Items in C4C/SPM v1.2m

Table 49.
Additional Non-measurement Items.

Item Text	Answer Choices (Multi-select)
If 17-BULL = 1 (I am being bullied now):	
Tell us about your experience being bullied…	
40-WHO Who is bullying you? [1]	Someone from outside the school (1) A student from another grade level (2) A student in my class (3) One of my teachers (4) Another adult in the school (5) Someone else (please tell us who) (6) [2]
41-WHY Why do you think the bully is picking on you? [1]	Because I'm disabled (1) Because of my racial/ethnic background (2) Because of my gender (3) Because they think I am gay (4) Because of my height or weight (5) Because of how I dress (6) Because of my school work (7) Because I don't fit in at school (8) Some other reason (please tell us why) (9) [2]
If 17-BULL = 2 (I was before but it has stopped now):	
Tell us about your experience being bullied in the past…	
50-WHO Who bullied you? [1]	Someone from outside the school (1) A student from another grade level (2) A student in my class (3) One of my teachers (4) Another adult in the school (5) Someone else (please tell us who) (6) [2]
51-WHY Why do you think the bully picked on you? [1]	Because I'm disabled (1) Because of my racial/ethnic background (2) Because of my gender (3) Because they think I am gay (4) Because of my height or weight (5) Because of how I dress (6) Because of my school work (7) Because I don't fit in at school (8) Some other reason (please tell us why) (9) [2]

Notes. [1]Respondent may select multiple options. [2]Free form response is available.

Cross-references of the Items Design and the Literature

Table 50.
Cross-reference between the Safety Construct and the Literature.

Subconstruct	Elementary Item	Middle School Item	Supporting Literature
Overall	19-SAFE How safe do you feel at your school overall?	22-SAFE How safe do you feel at your school overall?	Astor et al. (2009) Gregory et al. (2010) Gregory, Cornell, and Fan (2011)
Safety from Inside Threats	10B-B4 Before classes start for the day	12B-B4 Before classes start for the day	Astor et al. (2009) Pryce and Frederickson (2013) Veenstra et al. (2014) O'Connor et al. (2010)
	10C-CLS In class with my teacher there	12C-CLS In class with my teacher there	
	10D-LNCH At lunchtime	12D-LNCH At lunchtime	
	10E-RCSS At recess	-	
	10F-AFTR After classes end for the day	12F-AFTR After classes end for the day	
	13A-HURT They keep me safe from getting hurt	15A-HURT They keep me safe from getting hurt	Astor et al. (2009) Astor et al. (2010) Osher et al. (2010) Mitchell and Bradshaw (2013)
	13B-STFF They keep my stuff safe	15B-STFF They keep my stuff safe	
	13C-AFRD They help me not be afraid at school	15C-AFRD They help me not be afraid at school	
	13D-NICE They make school a nice place	15D-NICE They make school a nice place	
	13E-TCH They teach us about bullying	15E-TCH They teach us about bullying	
	14-TALK How often do you hear students talk about being mean to others?	16-TALK How often do you hear students talk about being mean to others?	Bouman et al. (2012) Olthof et al. (2011) Goldweber et al. (2013) Estell et al. (2008) Waasdorp and Bradshaw (2011)
	-	17-DRAM How often do you see drama happening at your school?	
	15-SEE How often do you see people bullying others at your school?	18-SEE How often do you see people bullying others at your school?	
	16-BULL Have you ever been bullied at school?	19-BULL Have you ever been bullied at school?	*See Table 52.*
	17A-STOP They stop the bullying	20A-STOP They stop the bullying	Henry et al. (2011) Frey et al. (2009) Bergsmann et al. (2013)
	17B-HELP They help the person who is hurt or upset	20B-HELP They help the person who is hurt or upset	
	17C-PUN They punish the bully	20C-PUN They punish the bully	
	17D-TALK They talk to the class about bullying	20D-TALK They talk to the class about bullying	

	18A-HELP They help the person who is hurt	21A-HELP They help the person who is hurt	Eyman and Cohen (2009) Batanova et al. (2014) Nickerson et al. (2008) Syvertsen et al. (2009)
	18B-STOP They tell the bully to stop	21B-STOP They tell the bully to stop	
	18C-TELL They tell an adult what happened	21C-TELL They tell an adult what happened	
Safe Built Environment	11A-CLS In my classroom	13A-CLS In my classroom	Astor et al. (2009) Colvin et al. (1997) Astor et al. (2001) Astor and Meyer (2001) McCurdy et al. (2009) Davies et al. (2013) Mayer and Furlong (2010)
	11B-HALL Walking in the hallways	13B-HALL Walking in the hallways	
	11C-BATH Going to the bathroom	13C-BATH Going to the bathroom	
	11D-PLGR Outside in the playground	13D-PLGR Outside in the playground	
	11E-STRS Walking up or down the stairs	13E-STRS Walking up or down the stairs	
	11F-PLOT Outside in the parking lot	13F-PLOT Outside in the parking lot	
	11G-LRM In the lunch room	13G-LRM In the lunch room	
Security from Outside Threats	10A-TRAV Traveling to and from school	12A-TRAV Traveling to and from school	Astor et al. (2009) Mayer and Furlong (2010) O'Toole (2000) Vossekuil et al. (2002) Borum et al. (2010) Cornell and Williams (2012) Miller (2012)
	12A-LOCK The outside doors are locked	14A-LOCK The outside doors are locked	
	12B-STKR Visitors have to wear stickers	14B-STKR Visitors have to wear stickers	
	12C-BUZZ People have to buzz in	14C-BUZZ People have to buzz in	
	12D-CHK The security guards check people	14D-CHK The security guards check people	

Table 51.
Supplemental Items: Personal Experience of Bullying (Middle School).

Being bullied now	**Bullied in the past, but stopped now**	**Supporting Literature**	
40-WHO Who is bullying you?	50-WHO Who bullied you?	Wang et al. (2012) Jutengren et al. (2011) Nansel et al. (2003) Juvonen et al. (2003) Woods and Wolke (2004) Little (2002) Norwich and Kelly (2004)	Marini et al. (2001) Janssen et al. (2004) Hoover et al. (1993) Olweus (1995) Robinson and Espelage (2011) Swearer et al. (2010) Scherr (2012)
41-WHY Why do you think the bully is picking on you?	51-WHY Why do you think the bully picked on you?		

Table 52.
Cross-reference between the Community Construct and the Literature.

Subconstruct	Elementary Item	Middle School Item	Supporting Literature
Overall	9-LIKE Overall, how much do you like your school?	11-LIKE Overall, how much do you like your school?	Astor et al. (2009) Gregory et al. (2010)
Strong Family Support	6A-HWK They help me with my homework	8A-HWK They help me with my homework	Nickerson et al. (2008) Adamski et al. (2013) Lawson and Alameda-Lawson (2011) Patall et al. (2008) Chan et al. (2013)
	6B-RPT They read my report cards	8B-RPT They read my report cards	
	6C-WKHD They encourage me to work hard	8C-WKHD They encourage me to work hard	
	6D-TTCH They talk to my teachers about me	8D-TTCH They talk to my teachers about me	
	6E-EVNT They come to school events	8E-EVNT They come to school events	
	-	8F-PTO They are involved in a parents' organization	
Student Cohesiveness	7A-FRND My closest school friends help me	9A-FRND My closest school friends help me	Martin and Dowson (2009) Ciani et al. (2010) Chapman et al. (2011) Gini et al. (2008) Loukas et al. (2012); Niehaus et al. (2012) Wang and Holcombe (2010)
	7B-CLS Other students in my class help me	9B-CLS Other students in my class help me	
	7C-STU Students in other grades help me	9C-STU Students in other grades help me	
	8A-KNSS Students are kind to other students	10A-KNSS Students are kind to other students	
	8B-RSSS Students are respectful to other students	10B-RSSS Students are respectful to other students	
	8E-KNST Students are kind to teachers	10E-KNST Students are kind to teachers	
	8F-RSST Students are respectful to teachers	10F-RSST Students are respectful to teachers	
Supportive School Staff	7D-TCHR My teachers help me	9D-TCHR My teachers help me	Hand (2009) Danielsen et al. (2010) Givens Rolland (2012) Bishop et al. (2013) Roorda et al. (2011) Rudasill et al. (2010) Turner et al. (2013)
	7E-ADLT Other adults in my school help me	9E-ADLT Other adults in my school help me	
	8C-KNTS Teachers are kind to students	10C-KNTS Teachers are kind to students	
	8D-RSTS Teachers are respectful to students	10D-RSTS Teachers are respectful to students	
Respect for Diversity	1-GR What grade are you in?	1-GR What grade are you in?	Monsen et al. (2013) Peter and Dalbert (2010) Goldsmith (2010)
	2-GNDR Are you a girl or a boy?	2-GNDR Are you a girl or a boy?	
	3-HISP Are you Hispanic/Latino?	3-HISP Are you Hispanic/Latino?	
	4-ENGL Do you usually speak English at home?	4-ENGL Do you usually speak English at home?	
	5-RACE What is your racial identity?	5-RACE What is your racial identity?	

Table 53.
Cross-reference between the Learning Construct and the Literature.

Subconstruct	Elementary Item	Middle School Item	Supporting Literature
High Expectations of All Students	-	23-FGRD What grade do you usually earn in [F] class?	Rubie-Davies et al. (2014) Corpus et al. (2009) Mainhard et al. (2011) Curby et al. (2013) Spearman and Watt (2013) McKown and Weinstein (2008)
	-	30-LGRD What grade do you usually earn in [LF] class?	
	20-EXP What does your teacher expect from you?	24-FEXP What does your [F] teacher expect from you?	
	-	31-LEXP What does your [LF] teacher expect from you?	
	22H-INSP Inspires me to do better	26G-FINS Inspires me to do better	
	-	33G-LINS Inspires me to do better	
Effective Teaching in All Classes	22A-DIR My teacher gives clear directions	26A-FDIR My teacher gives clear directions	Connor et al. (2009) Connor et al. (2014) Givens Rolland (2012) Marzano et al. (2003) Goetz et al. (2013)
	-	33A-LDIR My teacher gives clear directions	
	22G-TIME Gives me time to finish my work	26F-FTIM Gives me time to finish my work	
	-	33F-LTIM Gives me time to finish my work	
Respect for Diverse Learning Styles	22D-ENHL Encourages me to help others	26D-FENH Encourages me to help others	Brock et al. (2008) Rimm-Kaufman et al. (2014) Reis et al. (2010) Reznitskaya et al. (2012) Wirkala and Kuhn (2011)
	-	33D-LENH Encourages me to help others	
	23C-DIFF I get to do different things	27C-FDIF I get to do different things	
	-	34C-LDIF I get to do different things	
	23E-ACT Involves different kinds of activities	27E-FACT Involves different kinds of activities	
	-	34E-LACT Involves different kinds of activities	
	23F-TEAM Makes me work in teams with others	27F-FTM Makes me work in teams with others	
	-	34F-LTM Makes me work in teams with others	
Challenging Curriculum	21-DIFF How difficult is the work you do in class?	25-FDIF How difficult is the work you do in [F] class?	Cooper (2013) Martin and Dowson (2009) Plank and Condliffe (2013) Goetz et al. (2013)
	-	32-LDIF How difficult is the work you do in [LF] class?	
	25-LRN How often do you feel you learn new things in class?	29-FLRN Overall, how much do you feel you are learning in [F] class?	
	-	36-LLRN Overall, how much do you feel you are learning in [LF] class?	

Note. [F] respondent selected favorite subject. [LF] respondent selected least favorite subject.

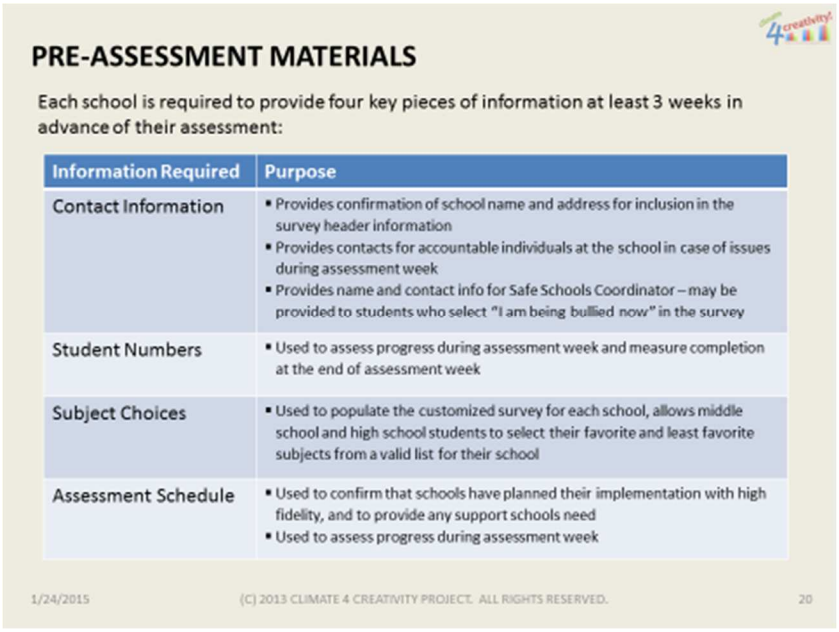
Table 54.
Cross-reference between the Creativity Construct and the Literature.

Subconstruct	Elementary Item	Middle School Item	Supporting Literature
Creative Focus & Flow	23A-FUN The work I do in class is fun	27A-FFUN The work I do in class is fun	Csikszentmihalyi (1996) Amabile et al. (2002) Mainemelis (2002) Antes and Mumford (2009) Newton (2013) Cheng (2011)
	-	34A-LFUN The work I do in class is fun	
	23D-TIME Makes me lose track of time	27D-FTIM Makes me lose track of time	
	-	34D-LTIM Makes me lose track of time	
Creative Freedom	22C-RETH Lets me rethink or reframe class problems	26C-FRET Lets me rethink or reframe class problems	Csikszentmihalyi (1996) Williams and Yang (2010) Agogué et al. (2014) Ciani et al. (2010) Abrami et al. (2008) Mathews and Lowe (2011)
	-	33C-LRET Lets me rethink or reframe class problems	
	22F-INV Lets me invent my own ways to solve problems	26H-FINV Lets me invent my own ways to solve problems	
	-	33H-LINV Lets me invent my own ways to solve problems	
Creative Challenge	23B-THNK Makes me think about stuff	27B-FTHK Makes me think about stuff	Csikszentmihalyi (1996) Davies et al. (2013) Cheng (2011)
	-	34B-LTHK Makes me think about stuff	
	23G-CREA Includes being creative	27G-FCRE Includes being creative	
	-	34G-LCRE Includes being creative	
Support for Creativity	22B-ENCR Encourages me to be creative	26B-FENC Encourages me to be creative	DiLiello et al. (2011) Amabile et al. (2002) Kim (2008) Peng et al. (2013) Runco (2007) Lassig (2013)
	-	33B-LENC Encourages me to be creative	
	22E-RECR Recognizes me for being creative	26E-FREC Recognizes me for being creative	
	-	33E-LREC Recognizes me for being creative	
	24-CRE How important do you think it is to be creative in class?	28-FCRE How important do you think it is to be creative in [F] class?	
	-	35-LCRE How important do you think it is to be creative in [LF] class?	

Note. [F] respondent selected favorite subject. [LF] respondent selected least favorite subject.

Appendix F – Climate4Creativity v1.2 Administration Instructions

Pre-assessment Instruction Materials



PRE-ASSESSMENT MATERIALS

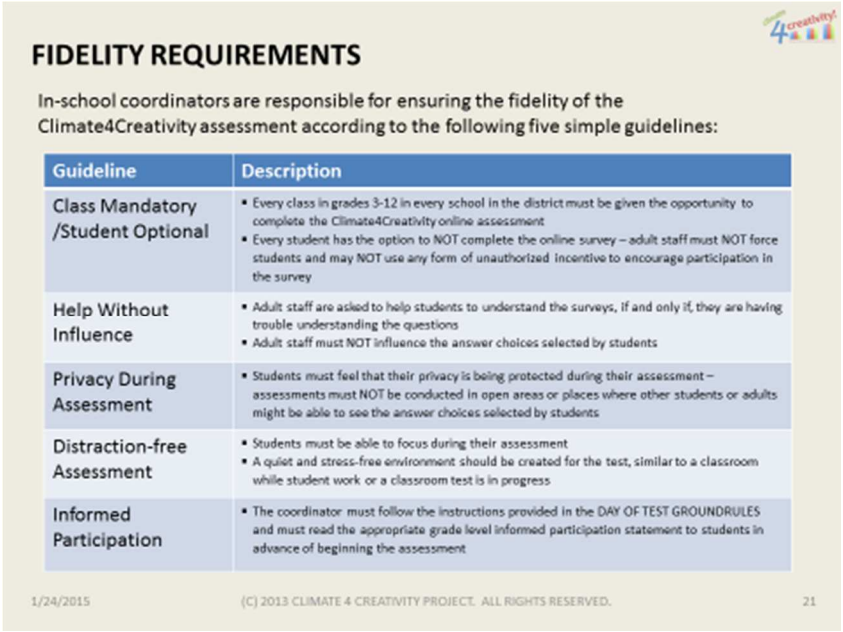
Each school is required to provide four key pieces of information at least 3 weeks in advance of their assessment:

Information Required	Purpose
Contact Information	▪ Provides confirmation of school name and address for inclusion in the survey header information ▪ Provides contacts for accountable individuals at the school in case of issues during assessment week ▪ Provides name and contact info for Safe Schools Coordinator – may be provided to students who select "I am being bullied now" in the survey
Student Numbers	▪ Used to assess progress during assessment week and measure completion at the end of assessment week
Subject Choices	▪ Used to populate the customized survey for each school, allows middle school and high school students to select their favorite and least favorite subjects from a valid list for their school
Assessment Schedule	▪ Used to confirm that schools have planned their implementation with high fidelity, and to provide any support schools need ▪ Used to assess progress during assessment week

1/24/2015 (C) 2013 CLIMATE 4 CREATIVITY PROJECT. ALL RIGHTS RESERVED. 20

Figure 55. Pre-assessment Instruction Materials.

Fidelity Requirements



FIDELITY REQUIREMENTS

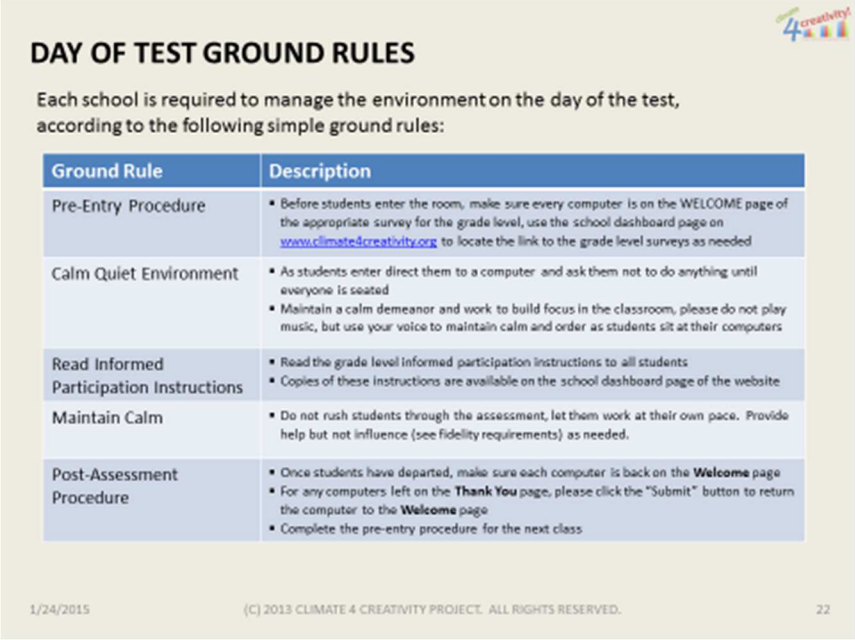
In-school coordinators are responsible for ensuring the fidelity of the Climate4Creativity assessment according to the following five simple guidelines:

Guideline	Description
Class Mandatory /Student Optional	▪ Every class in grades 3-12 in every school in the district must be given the opportunity to complete the Climate4Creativity online assessment ▪ Every student has the option to NOT complete the online survey – adult staff must NOT force students and may NOT use any form of unauthorized incentive to encourage participation in the survey
Help Without Influence	▪ Adult staff are asked to help students to understand the surveys, if and only if, they are having trouble understanding the questions ▪ Adult staff must NOT influence the answer choices selected by students
Privacy During Assessment	▪ Students must feel that their privacy is being protected during their assessment – assessments must NOT be conducted in open areas or places where other students or adults might be able to see the answer choices selected by students
Distraction-free Assessment	▪ Students must be able to focus during their assessment ▪ A quiet and stress-free environment should be created for the test, similar to a classroom while student work or a classroom test is in progress
Informed Participation	▪ The coordinator must follow the instructions provided in the DAY OF TEST GROUNDRULES and must read the appropriate grade level informed participation statement to students in advance of beginning the assessment

1/24/2015 (C) 2013 CLIMATE 4 CREATIVITY PROJECT. ALL RIGHTS RESERVED. 21

Figure 56. Fidelity Requirements.

Day of Test Ground Rules



DAY OF TEST GROUND RULES

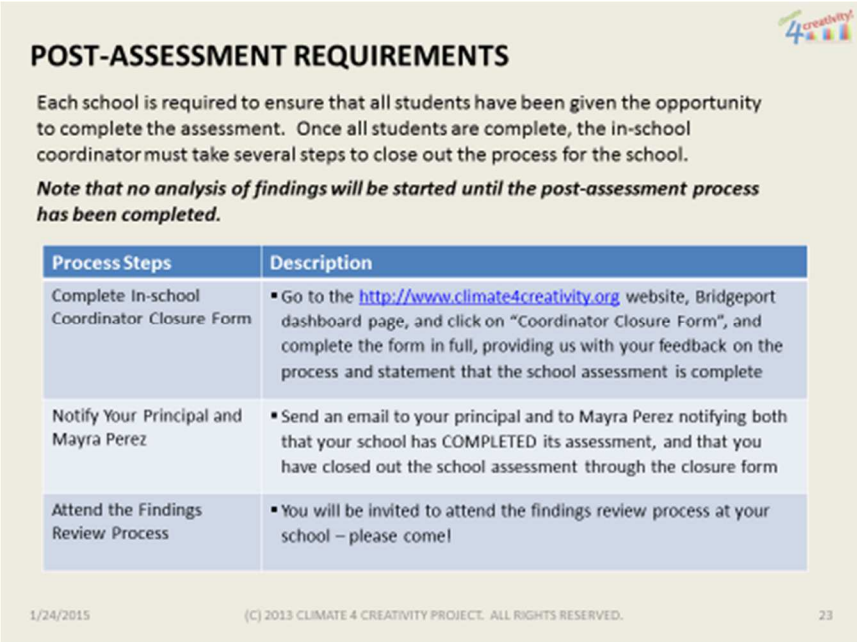
Each school is required to manage the environment on the day of the test, according to the following simple ground rules:

Ground Rule	Description
Pre-Entry Procedure	• Before students enter the room, make sure every computer is on the WELCOME page of the appropriate survey for the grade level, use the school dashboard page on www.climate4creativity.org to locate the link to the grade level surveys as needed
Calm Quiet Environment	• As students enter direct them to a computer and ask them not to do anything until everyone is seated • Maintain a calm demeanor and work to build focus in the classroom, please do not play music, but use your voice to maintain calm and order as students sit at their computers
Read Informed Participation Instructions	• Read the grade level informed participation instructions to all students • Copies of these instructions are available on the school dashboard page of the website
Maintain Calm	• Do not rush students through the assessment, let them work at their own pace. Provide help but not influence (see fidelity requirements) as needed.
Post-Assessment Procedure	• Once students have departed, make sure each computer is back on the **Welcome** page • For any computers left on the **Thank You** page, please click the "Submit" button to return the computer to the **Welcome** page • Complete the pre-entry procedure for the next class

1/24/2015 (C) 2013 CLIMATE 4 CREATIVITY PROJECT. ALL RIGHTS RESERVED. 22

Figure 57. Day of Test Ground Rules.

Post-Assessment Requirements



POST-ASSESSMENT REQUIREMENTS

Each school is required to ensure that all students have been given the opportunity to complete the assessment. Once all students are complete, the in-school coordinator must take several steps to close out the process for the school.

Note that no analysis of findings will be started until the post-assessment process has been completed.

Process Steps	Description
Complete In-school Coordinator Closure Form	• Go to the http://www.climate4creativity.org website, Bridgeport dashboard page, and click on "Coordinator Closure Form", and complete the form in full, providing us with your feedback on the process and statement that the school assessment is complete
Notify Your Principal and Mayra Perez	• Send an email to your principal and to Mayra Perez notifying both that your school has COMPLETED its assessment, and that you have closed out the school assessment through the closure form
Attend the Findings Review Process	• You will be invited to attend the findings review process at your school – please come!

1/24/2015 (C) 2013 CLIMATE 4 CREATIVITY PROJECT. ALL RIGHTS RESERVED. 23

Figure 58. Post-Assessment Requirements.

Administration Instructions (Script) Elementary School

Once all students are seated at computers, read the following informed participation instructions:

> *Please don't start or press any keys or move the mouse until I tell you to begin. Listen carefully to these instructions.*
>
> *You are going to be taking a survey on the computer. The survey will ask you some questions, and you will need to choose the answers by clicking on the buttons with the mouse. When you finish a page, click on the NEXT button at the bottom of the page, to move to the next page. If you get stuck or need help during the survey, please raise your hand and I will come over and try to help.*
>
> *Now, you should see a picture of Mr. P on the first page of the survey. Look at the screen now and please raise your hand if you don't see a picture of Mr. P.*

If any students raise their hands, check that they are on the Welcome page. If they are not, reset their computer so that are on the welcome page.

> *OK, now, let's read the instructions on the screen together: Hi! My name is Mr. P and I'm a teacher at a school in New Haven. I work with your school to help make it a safer and more friendly place. I need your help. The computer's going to ask you some questions about your school. I promise I won't tell your teacher what you said, so it is OK to say what you really think about your school.*
>
> *There's quite a few questions, but you don't have to answer them all. It would be great if you could answer as many as you can, but you can just leave them blank if you don't want to answer them. You can just click NEXT to go to the next page of questions.*
>
> *I'll come back at the end to let you know that you finished all the questions. Thank you for helping me make your school a safer and more friendly place.*
>
> *Now… Please click NEXT to start the questions. Don't forget, if you get stuck, please raise your hand and I will come over and try to help. You may begin.*

Now relax and wait for the students to complete or raise their hands for assistance!

Administration Instructions (Script) Middle School

Once all students are seated at computers, read the following informed participation instructions:

> *Please don't start using the computers until I tell you to begin. Listen carefully to these instructions.*
>
> *You are going to be taking a survey on the computer. The survey will ask you some questions, and you will need to choose the answers by clicking on the buttons with the mouse. When you finish a page, click on the NEXT button at the bottom of the page, to move to the next page. If you get stuck or need help during the survey, please raise your hand and I will come over and try to help.*
>
> *Now, you should be on the Welcome page of the survey, please look at the screen and raise your hand now if you are not on the Welcome page.*

If any students raise their hands, check that they are on the Welcome page. If they are not, reset their computer so that are on the welcome page.

> *OK, now, let's read the instructions on the screen together: Hi! My name is Mr. P and I'm a professor at a college in New Haven. I am working with your school district to help make your school a safer and more friendly place.*
>
> *To do this, I need your help. I'm going to ask you some questions about your school. Your answers will be kept confidential and I won't tell your teachers what you said, so it is OK to say what you really think about your school. It is really important that we know how things are in your school so we can help make them better for everyone.*
>
> *There's quite a few questions, but you don't have to answer them all. It would be great if you could answer as many as you can, but you can just leave them blank if you don't want to answer any of them. Some questions have * next to them - these ones are required. The survey won't move on if you leave these ones blank, so please do answer them.*
>
> *Thank you for helping make your school a safer and more friendly place.*
>
> *Now… Please click NEXT to start the questions. Don't forget, if you get stuck, please raise your hand and I will come over and try to help. You may begin.*

Now relax and wait for the students to complete or raise their hands for assistance!

Appendix G – Informed Legal Consent Materials

Study I – Expert Panel

Campus Site Approval – Sacred Heart University

Received by Email, July 5, 2012

Dear James:
Please consider our offer to host the July 10, 2012 School Climate Content Validity Panel in the Sacred Heart University Cambridge Commons from 5-9 PM. We look forward to having you conduct your work in the Commons in order to support the final survey instrument for Bridgeport Public Schools.

Jacqueline Kelleher

Jacqueline Kelleher, MA, Ph.D.
Assistant Professor
Isabelle Farrington School of Education
Sacred Heart University

Tel: 203-365-4472
Fax:203-365-7513
kelleherj@sacredheart.edu
http://www.sacredheart.edu/pages/33114_jacqueline_kelleher_ph_d_.cfm

Individual Participant Informed Consent

Introduction

You are being invited to take part in a research study. The information in this form is provided to help you decide whether or not to take part in this study. Study personnel will be available to answer your questions and provide additional information. If you decide to take part in the study, you will be asked to sign this consent form. A copy of this form will be given to you. If needed, study personnel will be available at the event to answer your questions and provide additional information.

What is the purpose of this research study?

This study is part of a research project designed to explore the relationships between school climate in a public school and the environment for teaching and learning creativity in that public school. The research project is designed to provide input to the following questions: What is the value of creative thinking to the sustainability of U.S. society? How might U.S. public school leaders create an environment in which creative thinking is treated as a foundational component of student learning outcomes, on a par with reading, writing, and numeracy? What are the enablers & barriers to creating a climate suited to teaching & learning creative thinking in U.S. public schools? What would this "climate for creativity" feel like? How could it be effectively measured? What process and product goals should be established to enable creation of this climate?

Why are you being asked to participate?

You are being invited because of your expertise and experience in one or more of four key disciplines required for this study:

- *Teaching (school teacher, special ed. teacher, school administrator, professor)*
- *Creativity (musician, artist, writer, creative business professional, or other creative disciplines)*
- *Student Support (school counselor, social worker, or other support disciplines)*
- *Survey Design (research, business, or academic survey expertise).*

How many people will be asked to participate in this study?

Approximately 25 persons will be asked to participate in this study.

What will happen during this study?

On arrival at the panel venue, you will be asked to participate in five (5) activities. A detailed agenda will be provided to you on arrival at the venue.

1. *The Principal Investigator will explain the purpose and conduct of the study and provide study materials to you. If you did not complete the online version of this form and the demographic information form in advance of the study, you will be asked to complete these forms during this discussion.*
2. *You will be asked to complete one "reference instrument". Two different instruments are being used for reference purposes and you will be randomly assigned an instrument to complete by the Principal Investigator.*

3. *You will be asked to complete the candidate items in the new instrument, along with an item assessment survey, scoring the items themselves on a variety of parameters.*
4. *You will be asked to participate in a facilitated discussion about the instrument.*
5. *You will be asked to complete a brief feedback form covering the event itself.*

Pizza and beverages will be provided midway through the event (following completion of the reference instrument), and an opportunity to meet other participants and engage in informal "table" discussions will commence at that time.

All study materials will be collected and placed in sealed envelopes at the end of the study.

How long will I be in this study?
About four (4) hours of your time is needed to complete this study.

Are there any risks to me?
The things that you will be doing involve very low risk. You may feel that some questions we ask will be stressful or upsetting. If this occurs you can stop participating immediately. If anything else during the meeting feels too risky to you, you can stop participating immediately.

Are there any benefits to me?
You will not receive any financial benefit or compensation from taking part in this study. This study is, however, the first phase of a project that we expect to have important repercussions in the Bridgeport, Connecticut public school system. Your participation will positively influence the quality of this upcoming work. You may derive professional benefit from participating in the instrument evaluation and discussion with other professionals.

Will there be any costs to me?
Aside from your time, there are no costs for taking part in the study.

Will I be paid to participate in the study?
No. You will not be paid for your participation. Food and beverages will be provided.

Will video or audio recordings be made of me during the study?
Provided no attendees object, the facilitated closing discussion will be audio taped in order to ensure accuracy of recording of the discussion. A transcript of the recording will be prepared following the panel, and the original recording will be destroyed. Names will be removed from the transcript and replaced with participant numbers.

Note: A separate consent is required for audio recording. Please check the consent box on the signature page of this document to give your consent for audio recording.

Will the information that is obtained from me be kept confidential?
Clearly, since you will be participating in a panel event, other participants, some of whom may be known to you, will be able to see you at the event. We will ask that participants only use first names in their conversations. You are, of course, at liberty to discuss the event with other participants. The only other person who will know that you participated in this study will be the Principal Investigator: James Patsalides.

Once the event has concluded, your records will be confidential. You will not be identified in any reports or publications resulting from the study. It is possible that representatives of the Federal Government or some other group that supports the research study will want to come to Prescott College to review the study. If that occurs, a copy of the information may be provided to them but your name will be removed before any information is released.

May I change my mind about participating?
Your participation in this study is voluntary. You may decide to not begin or to stop the study at any time. Any new information discovered about the research during the study will be provided to you.

Whom can I contact for additional information?
You can obtain further information about the research or voice concerns or complaints about the research by contacting the Principal Investigator, James Patsalides on (860) 830-5791 or by email at jpatsalides@prescott.edu.

If you have questions concerning your rights as a research participant, have general questions, concerns or complaints or would like to give input about the research and can't reach the researcher, or want to talk to someone other than the researcher, you may contact the Human Subjects Committee Chairperson at Prescott College, Dr. Noël Cox Caniglia at (928) 350-3201 or ncaniglia@prescott.edu.

Your Signature
By signing this form, I affirm that I have read the information contained in the form, that the study has been explained to me, that my questions have been answered and that I agree to take part in this study. I do not give up any of my legal rights by signing this form.

> ***Consent for Audio Recording***
> *Do you consent to the use of audio recording in this study, according to the information provided in the form? I understand that I am not obligated to provide consent for the use of audio recording (☑ please check one).*
>
> *☐ I consent to the use of audio recording*
> *☐ I do not consent to the use of audio recording.*

_____
Name (Printed)

_____ _____
Participant's Signature *Date signed*

Statement by person obtaining consent
I certify that I have explained the research study to the person who has agreed to participate, and that he or she has been informed of the purpose, the procedures, the possible risks and potential benefits associated with participation in this study. Any questions raised have been answered to the participant's satisfaction.

_James Patsalides_____
Name of study personnel

_____ _____
Study personnel Signature *Date signed*

Study II – Cognitive Lab

School Site Approval – Johnson Elementary School

January 25, 2013

James Patsalides
55 New Street
Manchester, CT 06040

Dear Mr. Patsalides,

I have reviewed your request regarding your study and am pleased to support your research project entitled, "The Climate for Creativity Project– Cognitive Interview Process". Your request to use the Geraldine W. Johnson School as a research site is granted. The research will include 30-40 minute, one-on-one interviews with 20-25 of our students, conducted during the school day. We understand that you will be onsite during completion of this research work at our site and will personally conduct the interview process with our students. A member of the District PBIS Team will be present throughout each interview.

Please note that this authorization covers the month of February 2013 only. We look forward to working with you on this project.

Sincerely,

Marlene A. Roberts

Marlene A. Roberts
Principal

School Site Approval Extension – Due to Bad Weather

February 25, 2013

James Patsalides
55 New Street
Manchester, CT 06040

Dear Mr. Patsalides,

In light of the recent interruption of the school schedule due to the snow storm and your team's inability to move forward with The Climate for Creativity Project–Cognitive Interview Process, I am granting you an extension to cover the month of March 2013 so the student interviews can be conducted. This will apply only to non-CMT testing times. We look forward to working with you on this project.

Sincerely,

Marlene A. Roberts

Marlene A. Roberts
Principal

Parent Invitation Letter

January 28, 2013

Dear Parent/Guardian:

I am excited to announce that Geraldine W. Johnson School has been selected to participate in an important research study as part of the Bridgeport School District's ongoing work to improve safe school climate in our schools. We have been asked to take part in the early stages of The Climate for Creativity Project– a significant research study being conducted by a professor from Albertus Magnus College in New Haven, designed to explore the relationships between safe school climate and the environment for teaching and learning in our schools.

We are inviting students in grades 3-8 to volunteer to take part in a 30-minute interview with one of the researchers on the project. Interviews will take place here at the school during the month of February 2013, and a member of the District PBIS Team will be present at all times during the interview process. The family of each student who participates in the interview process will receive a $10 Subway Restaurant Gift Card as a gift from the research team.

If you are willing to allow your child to take part, please read the attached pink sheet, complete the information on the back of the pink sheet, and return it to school before February 6, 2013. If you have any questions about the study, please email the researcher at the email address provided on the pink sheet.

Please make sure your child hands the pink sheet to his or her classroom teacher, so his or her information can be included in the list of potential interviewees– the research team will select a representative sample of 25 students from the list of all potential interviewees. Thank you.

Sincerely,

Marlene A. Roberts

Marlene A. Roberts
Principal

Parental Consent Form

Informed Consent
--- Please Return This Pink Sheet To School With Your Consent ---

The Climate for Creativity Project – Cognitive Interview Process

Your child is being invited to take part in a research study. The information on this pink sheet is to help you decide whether or not to let your child take part in the study.

What is the purpose of this research study?
The study will look at some questions from a survey to improve the quality of the questions on the survey. The survey will later be used to explore the connections between safe school climate and teaching and learning in a public school.

How many people will be asked to take part in this study?
All students in grades 3-8 at Geraldine W. Johnson School are being invited to take part. About 25 students will be selected from those whose parents agree to let them take part.

What will happen during this study?
Your child will be interviewed by a researcher. The interview process will take about 30 minutes and will take place on school premises during the normal school day.

Are there any risks to my child?
*This study involves very low risk. If your child feels uncomfortable during the interview, he or she can stop immediately by saying **STOP NOW** to the researcher. A member of the school's social work team will be present throughout the interview process.*

Are there any benefits or costs to me from this study?
If your child is selected and completes the interview process, a $10 Subway Restaurants gift card will be mailed to you at the address you provide on this form. Gift cards will be mailed March 1, 2013. Aside from your child's time, there are no costs for taking part in the study.

Will video or sound recordings be made of my child during the study?
Yes. A sound only (digital audio) recording will be made of the interview. No video or photographic images will be made during the interview.

Will the information that is obtained from my child be kept confidential?
Yes. All data collected through this process will be kept confidential and will not be shared outside the research team. Your child's name will be protected and maintained separately from the digital recording. All data will be kept locked away and will be password protected.

May I change my mind about participating?
*Your child's participation in this study is voluntary. Your child may decide to not begin or to stop the study at any time by saying **STOP NOW** to the researcher.*

Who can I contact for more information?
You can obtain more information about this research or voice concerns or complaints by contacting the researcher, James Patsalides, by email at jpatsalides@prescott.edu.

Who can I contact if I have concerns about this research?
If you have questions about your child's rights as a research participant, have general questions, concerns or complaints, or would like to give input about the research and can't reach the researcher, or if you want to talk to someone other than the researcher, you may contact the Human Subjects Committee Chairperson at Prescott College, Dr. Noël Cox Caniglia by email at ncaniglia@prescott.edu.

Your Signatures
I affirm that I have read the information contained in the form, that the study has been explained to me, that my questions have been answered and that I agree to allow my child to take part in this study. I do not give up any of my legal rights by signing this form.

Consent for Sound Recording
Do you consent to the use of sound recording in this study, according to the information provided in the form? I understand that I am not obligated to provide consent for the use of sound recording. ☑ *please check one.*

☐ *I agree to the use of sound recording*

☐ *I do not agree to the use of sound recording.*

_____ _____
Parent/guardian name (Printed) Child's name (Printed)

_____ _____ _____ CT _____
Parent/guardian address Apt# City State Zip

_____ _____
Parent/guardian signature Date signed

Statement by person obtaining consent
I certify that I have explained the research study to the person who has agreed to participate, and that he or she has been informed of the purpose, the procedures, the possible risks and potential benefits associated with participation in this study. Any questions raised have been answered to the participant's satisfaction.

_James Patsalides_____
Name of study personnel

_____ **_1/15/2013_____**
Study personnel signature

Verbal Assent Script (recorded on digital audio)

Hi. My name is Mr. P and I'm a teacher in New Haven. I'd like to ask you a few questions to help us do some research work for your school.

What's your name? _____ *And your grade?* _____

This is Ms. Perez. Do you know her? _____ *She's going to sit with us while we talk, and she can help you if you need anything while we talk. OK?* ____

I'm going to tell you a bit about what I'd like to talk with you about so that you can be certain you want to do it. You don't have to take part if you don't want to, just tell me now and you can go right back to your class and nobody will mind. Is that OK? ____

I am trying to make a survey to give to everyone in the Bridgeport schools, and I'd like to make the questions that we ask on the survey as good as I can, so that they're easier for all the students in Bridgeport to answer. I'd like to ask you some of the questions from the survey and get your ideas about how to improve them. I'm interested in your ideas because you're a student here at Johnson Elementary and I'll be asking about 25 students from your school to take part.

Do you have any questions about why we're doing this? ____

If you choose to take part, I will show you some lists and ask you some questions. It will take about half an hour of your time. Some of the questions are about bullying and school safety. You might feel a bit stressed or upset when I ask those questions. If that happens, just say STOP NOW and we'll stop asking you questions. If anything else during questions feels risky, just say STOP NOW and we'll stop asking you questions right away. Ms. Perez is right here to help you if you need her. Is that OK? ____

We will be making a sound recording of our interview, so that I can go back and listen later and make sure I understood exactly what you said. Is that OK with you? ____

We won't pass any of your answers to your teachers or your family. Of course, you can tell them about what we talk about, if you want. But I will keep our conversation just between us. Is that OK? ____

Remember, you don't have to take part. If you don't want to, or if you want to stop, just say STOP NOW, and we'll stop right away. Ms. Perez is here with us to make sure you have help if you need it. OK? ____

Do you have any questions for me before we begin? ____

Is it OK to begin now? ____

Study III – Pilot Field Test

School Site Approval – High Horizon's Magnet School

January 14, 2013

James Patsalides
55 New Street
Manchester CT 06040

Dear Mr. Patsalides:

I have reviewed your request regarding your study and am pleased to support your research project entitled, "The Climate for Creativity Project – Field Test". Your request to use High Horizons Magnet School as a research site is granted. The research will include completion of an online survey by our students during the school day, under the supervision of school staff.

Please note that this authorization covers the time period from February 4, 2013 to February 8, 2013 only. You are welcome to join us during completion of the research work at our site.

We look forward to working with you on this project.

Sincerely,

Dr. Melissa Jenkins
Principal

Participant Online Consent – Elementary School Grades 3-5

Presented on the first page of the online survey, respondent must click "Next" to continue.

Hi! My name is Mr. P and I'm a teacher at a school in New Haven. I work with your school to help make it a safer and more friendly place.

I need your help. The computer's going to ask you some questions about your school. I promise I won't tell your teacher what you said, so it is OK to say what you really think about your school.

There's quite a few questions, but you don't have to answer them all. It would be great if you could answer as many as you can, but you can just leave them blank if you don't want to answer them.

Look at the bottom of the page now. See how it shows you how many pages you've already done and how many are left to do? You can just click NEXT to go to the next page of questions.

I'll come back at the end to let you know that you finished all the questions. Thank you for helping me make your school a safer and more friendly place.

Please click NEXT to start the questions.

Participant Online Consent – Middle School Grades 6-8

Presented on the first page of the online survey, respondent must click "Next" to continue.

Hi! My name is Mr. Patsalides and I'm a professor at a college in New Haven. I am working with your school district to help make your school a safer and more friendly place.

To do this, I need your help. I'm going to ask you some questions about your school. Your answers will be kept confidential and I won't tell your teacher what you said, so it is OK to say what you really think about your school. It is really important that we know how things are in your school so we can help make them better for everyone.

There's quite a few questions, but you don't have to answer them all. It would be great if you could answer as many as you can, but you can just leave them blank if you don't want to answer any of them. Some questions have * next to them - these ones are required. The survey won't move on if you leave these ones blank, so please do answer them.

You'll notice that at the bottom of the page it shows you how much of this survey you have completed so far - right now it says 9% completed. You just click NEXT at the bottom to continue onto the next page of questions.

Thank you for helping make your school a safer and more friendly place.

Please click NEXT to start the questions.

Study IV – Calibration Study

Districtwide Approval Letter

BOARD OF EDUCATION

CITY HALL - 45 LYON TERRACE
BRIDGEPORT, CONNECTICUT 06604

PAUL G. VALLAS
Superintendent of Schools

MEMBERS OF THE BOARD

JACQUELINE KELLEHER
Chairperson

KENNETH MOALES, JR.
Vice Chairperson

HERNAN ILLINGWORTH
Secretary

MEMBERS OF THE BOARD cont.

BOBBY SIMMONS

SAUDA BARAKA

LETICIA COLON

MARIA PERIERA

THOMAS A. MULLIGAN, JR.

JOHN BAGLEY

2006 & 2007 BROAD FINALIST
"EXPECT GREAT THINGS!"

March 14, 2013

James Patsalides
55 New Street
Manchester CT 06040

Dear Mr. Patsalides:

I have reviewed your request regarding your study and am pleased to support your research project entitled, "The Climate for Creativity Project". Your request to use the Bridgeport School System as a research site is granted. The research will include completion of an electronic survey for all students in grades 3-12 enrolled in our school system, conducted during the school day under supervision of school staff located in our schools. We understand that you will be onsite periodically during completion of this research and will work directly with school staff responsible for each site to coordinate completion of research work at each site.

In addition, please note that your research will be used to satisfy certain statutory requirements for our schools, and as such we do not expect you to obtain individual parental and student consent for this assessment. We understand that all students in grades 3-12 will be invited to participate in the electronic survey, but some may choose not to participate. We take our commitment to protecting the rights of our students and parents very seriously, and we will ensure that adequate protections are in place for our students throughout this research process.

Please note that this authorization covers the time period from today through June 1, 2014 only. We look forward to working with you on this project.

Sincerely,

Sandra Kase

Sandra Kase, Ed.D.
Chief Administrative Officer
Bridgeport Public Schools

Participant Consent – Elementary School Grades 3-5 or 3-6

Presented on the first page of the online survey, respondent must click "Next" to continue.

Hi! My name is Mr. P and I'm a teacher at a school in New Haven. I work with your school to help make it a safer and more friendly place.

I need your help. The computer's going to ask you some questions about your school. I promise I won't tell your teacher what you said, so it is OK to say what you really think about your school.

There's quite a few questions, but you don't have to answer them all. It would be great if you could answer as many as you can, but you can just leave them blank if you don't want to answer them.

I'll come back at the end to let you know that you finished all the questions. Thank you for helping me make your school a safer and more friendly place.

Please click NEXT to start the questions.

Participant Consent – Middle School Grades 6-8

Presented on the first page of the online survey, respondent must click "Next" to continue.

Hi! My name is Mr. Patsalides and I'm a professor at a college in New Haven. I am working with your school district to help make your school a safer and more friendly place.

To do this, I need your help. I'm going to ask you some questions about your school. Your answers will be kept confidential and I won't tell your teachers what you said, so it is OK to say what you really think about your school. It is really important that we know how things are in your school so we can help make them better for everyone.

There's quite a few questions, but you don't have to answer them all. It would be great if you could answer as many as you can, but you can just leave them blank if you don't want to answer any of them. Some questions have * next to them - these ones are required. The survey won't move on if you leave these ones blank, so please do answer them.

You'll notice that at the bottom of the page it shows you how much of this survey you have completed so far - right now it says 9% completed. You just click NEXT at the bottom to continue onto the next page of questions.

Thank you for helping make your school a safer and more friendly place.

Please click NEXT to start the questions.

Appendix H – Empirical Fit Statistics for All Measures

Empirical fit statistics are shown in this appendix for each construct in both instruments. This appendix is included for the purposes of providing complete technical documentation on the instruments developed and tested in this study. Note that Winsteps *table 10.1* Item misfit tables are provided (Linacre, 2012) showing both infit and outfit mean-squares parameters. Detailed definitions of each item can be found in Appendix E using the item codes, shown on the right of each table row.

Empirical Fit Statistics: Elementary Instrument

Table 55.
Empirical Fit Statistics: Climate4Creativity Elementary Overall Measure.

```
TABLE 10.1 C4C/E Overall - All Schools          ZOU055WS.TXT  Apr 13 23:08 2015
INPUT: 4684 Student  78 Item  REPORTED: 4628 Student  58 Item  5 CATS WINSTEPS 3.81.0
-------------------------------------------------------------------------------
Student: REAL SEP.: 2.96  REL.: .90 ... Item: REAL SEP.: 27.01  REL.: 1.00
         Item STATISTICS:  MISFIT ORDER
-------------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL           MODEL|   INFIT  |  OUTFIT  |PTMEASURE-A|EXACT MATCH|         |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item    |
|------------------------------------+----------+----------+-----------+-----------+---------|
|    12  20277   4405    -.92     .02|1.63   9.9|1.62   9.9|A .25   .33| 66.3  65.5| 6C-WKHD |
|    17   9855   4295    1.24     .01|1.35   9.9|1.52   9.9|B .32   .51| 26.3  29.6| 7C-STU  |
|    54  18025   4363    -.23     .02|1.41   9.9|1.52   9.9|C .39   .43| 33.2  37.8| 17C-PUN |
|    13  15121   4375     .37     .01|1.30   9.9|1.42   9.9|D .27   .50| 27.1  30.0| 6D-TTCH |
|    49  11157   4537    1.11     .01|1.15   7.7|1.42   9.9|E .29   .52| 30.3  28.8| 14-TALK |
|    77  16085   4416     .23     .01|1.15   7.4|1.42   9.9|F .26   .48| 34.9  31.0| 24-CRE  |
|    15  16051   4455     .25     .01|1.27   9.9|1.38   9.9|G .29   .49| 28.7  30.7| 7A-FRND |
|    19  14489   4369     .47     .01|1.32   9.9|1.38   9.9|H .36   .51| 25.2  29.6| 7E-ADLT |
|    14  13061   4375     .72     .01|1.21   9.9|1.33   9.9|I .30   .52| 28.9  29.1| 6E-EVNT |
|    55  16932   4367     .02     .01|1.26   9.9|1.31   9.9|J .44   .46| 28.8  33.0| 17D-TALK|
|    50  13661   4542     .70     .01|1.15   8.1|1.27   9.9|K .35   .52| 29.8  29.1| 15-SEE  |
|    10  15556   4539     .39     .01|1.15   8.0|1.26   9.9|L .31   .50| 30.5  29.9| 6A-HWK  |
|    69  19454   4302    -.76     .02|1.25   7.5|1.03    .9|M .46   .35| 63.7  58.2| 22H-INSP|
|    38  13634   4367     .62     .01|1.17   9.2|1.21   9.9|N .42   .51| 29.4  29.3| 11F-PLOT|
|    46  17379   4419    -.03     .01|1.20   9.0|1.11   4.3|O .57   .46| 31.6  33.8| 13C-AFRD|
|    18  18764   4412    -.36     .02|1.10   4.0|1.16   5.2|P .39   .41| 40.2  41.8| 7D-TCHR |
|    57  16790   4357     .04     .01|1.16   7.3|1.11   4.4|Q .51   .47| 28.6  32.5| 18B-STOP|
|    67  16404   4330     .10     .01|1.16   7.5|1.15   6.0|R .48   .47| 29.1  32.0| 22F-INV |
|    58  18371   4395    -.28     .02|1.15   6.0|1.05   1.9|S .49   .42| 37.2  39.2| 18C-TELL|
|    53  19308   4383    -.57     .02|1.13   4.5|1.06   1.7|T .48   .38| 54.0  49.7| 17B-HELP|
|    23  19420   4350    -.66     .02|1.10   3.2| .97   -.8|U .48   .37| 58.3  53.7| 8D-RSTS |
|    52  17603   4448    -.05     .01|1.09   4.0|1.08   3.0|V .50   .45| 33.4  34.0| 17A-STOP|
|    63  19005   4388    -.46     .02|1.08   3.1| .94  -1.9|W .50   .40| 50.2  45.9| 22B-ENCR|
|    45  17054   4430     .05     .01|1.07   3.2|1.02   1.0|X .55   .47| 32.7  32.5| 13B-STFF|
|    65  17895   4367    -.19     .02|1.07   3.1| .99   -.5|Y .52   .44| 37.9  36.9| 22D-ENHL|
|    66  17559   4316    -.16     .02|1.07   2.9| .98   -.7|Z .52   .44| 37.5  36.0| 22E-RECR|
|    32  17733   4365    -.15     .02|1.06   2.5| .99   -.3|  .44   .44| 40.8  35.9| 10F-AFTR|
|    72  16934   4284    -.04     .01|1.05   2.5|1.06   2.5|  .47   .45| 34.2  33.8| 23C-DIFF|
|       BETTER FITTING OMITTED       +----------+----------+           |           |         |
|    76  17113   4260    -.10     .02| .98   -.7| .94  -2.5|z .52   .45| 34.0  34.9| 23G-CREA|
|    47  17594   4405    -.09     .01| .97  -1.6| .87  -5.4|y .61   .45| 35.0  34.6| 13D-NICE|
|    71  16752   4289     .00     .01| .97  -1.5| .95  -2.0|x .49   .46| 37.0  33.3| 23B-THNK|
|    16  13969   4415     .59     .01| .89  -6.4| .96  -2.3|w .40   .51| 35.4  29.3| 7B-CLS  |
|    22  19094   4397    -.48     .02| .96  -1.5| .89  -3.6|v .50   .39| 50.6  46.3| 8C-KNTS |
|    68  17790   4318    -.21     .02| .96  -1.8| .91  -3.3|u .51   .43| 38.4  37.5| 22G-TIME|
|    74  17335   4249    -.17     .02| .95  -2.2| .93  -2.8|t .50   .44| 39.4  36.3| 23E-ACT |
|    75  15910   4282     .17     .01| .88  -6.1| .94  -2.9|s .46   .48| 33.7  31.4| 23F-TEAM|
|    27  18480   4468    -.23     .02| .93  -3.3| .91  -3.6|r .42   .43| 46.0  38.0| 10A-TRAV|
|    31  16364   4337     .12     .01| .93  -3.4| .91  -4.2|q .47   .47| 36.0  31.6| 10E-RCSS|
|    37  17150   4388     .00     .01| .92  -3.8| .90  -4.1|p .47   .46| 41.3  33.3| 11E-STRS|
|    62  20012   4458    -.70     .02| .92  -2.7| .84  -4.9|o .48   .36| 60.3  55.5| 22A-DIR |
|    26  16062   4309     .16     .01| .87  -6.6| .91  -4.3|n .53   .48| 35.7  31.4| 9-LIKE  |
|    39  17767   4384    -.14     .02| .91  -4.0| .87  -5.3|m .50   .44| 44.5  35.8| 11G-LRM |
|    30  17232   4416     .00     .01| .90  -4.8| .89  -4.9|l .50   .46| 39.7  33.2| 10D-LNCH|
|    35  16861   4389     .06     .01| .89  -5.7| .90  -4.5|k .48   .47| 39.7  32.5| 11C-BATH|
|    33  20051   4508    -.63     .02| .88  -4.5| .80  -6.4|j .48   .37| 59.5  52.5| 11A-CLS |
|    36  16680   4393     .10     .01| .88  -6.1| .86  -6.4|i .48   .47| 38.8  31.9| 11D-PLGR|
|    25  15939   4367     .21     .01| .83  -9.7| .85  -7.0|h .47   .48| 35.1  30.9| 8F-RSST |
|    24  15840   4339     .22     .01| .79  -9.9| .83  -8.4|g .47   .48| 36.7  30.9| 8E-KNST |
|    34  17452   4437    -.03     .01| .79  -9.9| .79  -9.2|f .48   .46| 45.3  33.7| 11B-HALL|
|    59  17224   4443     .03     .01| .79  -9.9| .77  -9.9|e .59   .46| 41.5  33.1| 19-SAFE |
|    28  18587   4395    -.33     .02| .77  -9.9| .77  -8.6|d .45   .41| 50.8  41.0| 10B-B4  |
|    70  15616   4385     .29     .01| .74  -9.9| .75  -9.9|c .55   .49| 37.6  30.5| 23A-FUN |
|    21  14082   4417     .57     .01| .69  -9.9| .74  -9.9|b .50   .51| 40.1  29.4| 8B-RSSS |
|    20  14574   4489     .53     .01| .69  -9.9| .73  -9.9|a .52   .51| 40.1  29.5| 8A-KNSS |
|------------------------------------+----------+----------+-----------+-----------+---------|
| MEAN 16889.0 4388.8     .00     .02|1.03    .8|1.03    .0|           | 39.3  36.5|         |
| S.D.  2076.9   66.5     .43     .00| .18   6.6| .22   6.5|           |  9.9   8.4|         |
-------------------------------------------------------------------------------------------------
```

Table 56.
Empirical Fit Statistics: Climate4Creativity Elementary Safety Measure.

TABLE 10.1 C4C/E Safety - All Schools ZOU951WS.TXT Apr 13 23:17 2015
INPUT: 4684 Student 78 Item REPORTED: 4608 Student 27 Item 5 CATS WINSTEPS 3.81.0

Student: REAL SEP.: 2.35 REL.: .85 ... Item: REAL SEP.: 24.72 REL.: 1.00
Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASURE-A CORR.	PTMEASURE-A EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
49	11157	4537	1.25	.01	1.20	9.6	1.64	9.9	A .40	.60	32.7	32.2	14-TALK
54	18025	4363	-.21	.02	1.48	9.9	1.64	9.9	B .41	.48	38.1	43.1	17C-PUN
55	16932	4367	.06	.01	1.42	9.9	1.60	9.9	C .41	.52	32.2	38.2	17D-TALK
50	13661	4542	.81	.01	1.13	7.0	1.27	9.9	D .46	.59	32.0	30.6	15-SEE
46	17379	4419	.01	.02	1.23	9.8	1.14	4.9	E .55	.51	36.3	38.6	13C-AFRD
57	16790	4357	.09	.01	1.19	8.5	1.20	7.2	F .51	.52	32.3	37.1	18B-STOP
58	18371	4395	-.26	.02	1.20	7.5	1.13	4.0	G .48	.47	40.9	44.5	18C-TELL
38	13634	4367	.71	.01	1.13	6.7	1.19	8.3	H .51	.58	32.3	30.7	11F-PLOT
53	19308	4383	-.58	.02	1.18	6.1	1.07	2.1	I .47	.43	56.5	53.8	17B-HELP
52	17603	4448	-.01	.02	1.11	4.9	1.13	4.5	J .51	.51	37.8	39.3	17A-STOP
29	19985	4412	-.79	.02	1.09	2.6	.95	-1.3	K .43	.40	63.0	61.2	10C-CLS
45	17054	4430	.09	.01	1.08	3.7	1.04	1.4	L .56	.52	37.8	37.3	13B-STFF
56	17720	4382	-.11	.02	1.07	2.8	1.06	2.0	M .51	.50	39.3	41.4	18A-HELP
32	17733	4365	-.12	.02	1.03	1.3	.94	-2.0	N .50	.49	44.3	41.7	10F-AFTR
27	18480	4468	-.21	.02	.96	-1.8	.98	-.7	m .45	.48	48.1	43.1	10A-TRAV
44	18647	4511	-.21	.02	.98	-.8	.90	-3.6	l .56	.48	45.2	43.2	13A-HURT
47	17594	4405	-.05	.02	.97	-1.5	.88	-4.5	k .60	.50	41.6	40.3	13D-NICE
31	16364	4337	.17	.01	.90	-5.0	.89	-4.6	j .53	.53	40.0	36.4	10E-RCSS
33	20051	4508	-.64	.02	.90	-3.5	.82	-5.2	i .48	.42	61.3	56.1	11A-CLS
37	17150	4388	.04	.02	.89	-5.0	.86	-5.4	h .53	.51	46.5	38.4	11E-STRS
35	16861	4389	.10	.01	.85	-7.6	.83	-7.0	g .55	.52	44.4	37.3	11C-BATH
30	17232	4416	.04	.01	.84	-7.8	.81	-7.5	f .56	.51	44.4	38.3	10D-LNCH
39	17767	4384	-.11	.02	.84	-7.5	.77	-8.8	e .56	.49	50.0	41.6	11G-LRM
36	16680	4393	.15	.01	.82	-8.9	.80	-8.2	d .55	.53	43.9	36.5	11D-PLGR
28	18587	4395	-.32	.02	.78	-9.3	.80	-6.6	c .48	.46	52.5	46.9	10B-B4
34	17452	4437	.01	.02	.75	-9.9	.75	-9.9	b .55	.51	48.5	38.6	11B-HALL
59	17224	4443	.07	.01	.73	-9.9	.74	-9.9	a .61	.52	46.1	38.2	19-SAFE
MEAN	17238.6	4416.3	.00	.02	1.03	.4	1.03	-.4			43.3	40.9	
S.D.	1849.9	54.2	.41	.00	.19	6.9	.26	6.6			8.3	6.9	

Table 57.
Empirical Fit Statistics: Climate4Creativity Elementary Community Measure.

TABLE 10.1 C4C/E Community - All Schools ZOU064WS.TXT Apr 13 23:19 2015
INPUT: 4684 Student 78 Item REPORTED: 4625 Student 17 Item 3 CATS WINSTEPS 3.81.0

Student: REAL SEP.: 1.56 REL.: .71 ... Item: REAL SEP.: 37.73 REL.: 1.00
Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASURE-A CORR.	PTMEASURE-A EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
11	12876	4484	-2.89	.05	1.34	9.3	1.40	5.9	A .24	.29	88.5	87.5	6B-RPT
17	7306	4295	2.41	.03	1.24	9.9	1.24	9.9	B .46	.50	59.0	65.0	7C-STU
19	9374	4369	.66	.03	1.23	9.8	1.22	9.3	C .51	.48	61.9	68.8	7E-ADLT
12	12162	4405	-2.04	.04	1.22	9.2	1.22	5.2	D .33	.35	78.3	77.9	6C-WKHD
26	9826	4309	.15	.03	1.13	6.3	1.20	8.7	E .46	.47	63.5	66.9	9-LIKE
15	10036	4455	.25	.03	1.17	8.1	1.19	8.2	F .43	.47	63.2	67.3	7A-FRND
13	9702	4375	.38	.03	1.16	7.1	1.15	6.7	G .41	.47	64.8	67.9	6D-TTCH
14	8839	4375	1.16	.03	1.12	5.1	1.12	5.1	H .43	.49	67.2	69.4	6E-EVNT
10	10012	4539	.44	.03	1.04	2.1	1.04	1.9	I .44	.48	68.4	68.1	6A-HWK
23	11625	4350	-1.54	.03	1.01	.6	.94	-2.1	h .47	.39	74.5	71.9	8D-RSTS
18	11311	4412	-1.03	.03	.96	-2.1	.95	-2.0	g .49	.42	72.2	67.1	7D-TCHR
22	11417	4397	-1.18	.03	.94	-3.7	.90	-3.8	f .50	.41	73.2	68.0	8C-KNTS
25	9877	4367	.20	.03	.82	-9.5	.81	-9.3	e .50	.47	72.6	67.2	8F-RSST
24	9781	4339	.24	.03	.78	-9.9	.78	-9.9	d .49	.47	73.8	67.3	8E-KNST
16	9125	4415	.98	.03	.76	-9.9	.76	-9.9	c .50	.48	77.1	69.4	7B-CLS
20	9424	4489	.86	.03	.68	-9.9	.67	-9.9	b .51	.49	79.9	69.2	8A-KNSS
21	9165	4417	.95	.03	.65	-9.9	.65	-9.9	a .51	.49	80.9	69.4	8B-RSSS
MEAN	10109.3	4399.5	.00	.03	1.02	.8	1.01	.2			71.7	69.9	
S.D.	1330.3	62.7	1.28	.00	.21	7.8	.22	7.5			7.7	5.1	

Table 58.
Empirical Fit Statistics: Climate4Creativity Elementary Higher Order Learning Measure.

```
TABLE 10.1 C4C/E Learning/Creativity - All Schoo ZOU591WS.TXT  Apr 13 23:21 2015
INPUT: 4684 Student  78 Item  REPORTED: 4533 Student  15 Item  3 CATS WINSTEPS 3.81.0
-------------------------------------------------------------------------------
Student: REAL SEP.: 1.99  REL.: .80 ... Item: REAL SEP.: 19.98  REL.: 1.00
         Item STATISTICS:  MISFIT ORDER
-------------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL           MODEL|   INFIT  |  OUTFIT  |PTMEASURE-A|EXACT MATCH|         |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item    |
|------------------------------------+----------+----------+-----------+-----------+---------|
|    77   9715   4416    1.12     .03|1.41   9.9|1.57   9.9|A .38   .63| 62.3  68.6| 24-CRE  |
|    67  10140   4330     .48     .03|1.11   5.2|1.07   3.1|B .63   .62| 67.3  68.9| 22F-INV |
|    72  10314   4284     .18     .03|1.08   3.9|1.05   2.1|C .59   .61| 69.5  69.3| 23C-DIFF|
|    65  10842   4367    -.18     .03|1.03   1.3| .99   -.4|D .62   .59| 71.9  70.3| 22D-ENHL|
|    71  10098   4289     .42     .03|1.00    .2| .99   -.4|E .60   .61| 70.8  69.1| 23B-THNK|
|    64  10857   4356    -.22     .03| .99   -.5| .94  -2.2|F .62   .59| 73.2  70.5| 22C-RETH|
|    62  11834   4458   -1.07     .04| .97  -1.5| .94  -1.5|G .56   .55| 75.7  75.0| 22A-DIR |
|    66  10600   4316    -.04     .03| .97  -1.3| .91  -3.8|H .66   .60| 71.8  69.8| 22E-RECR|
|    68  10749   4318    -.20     .03| .97  -1.3| .97  -1.1|g .60   .59| 71.9  70.5| 22G-TIME|
|    69  11666   4302   -1.42     .04| .97  -1.2| .86  -3.2|f .59   .53| 80.8  77.4| 22H-INSP|
|    63  11364   4388    -.72     .03| .94  -3.1| .85  -4.8|e .63   .57| 76.8  72.8| 22B-ENCR|
|    74  10457   4249    -.07     .03| .90  -5.0| .84  -6.3|d .63   .60| 74.0  69.9| 23E-ACT |
|    70   9719   4385    1.03     .03| .89  -5.5| .89  -5.3|c .59   .63| 72.2  68.9| 23A-FUN |
|    75   9844   4282     .68     .03| .88  -6.1| .87  -6.0|b .60   .62| 73.4  69.0| 23F-TEAM|
|    76  10418   4260     .00     .03| .87  -6.8| .80  -8.3|a .67   .60| 73.6  69.7| 23G-CREA|
|------------------------------------+----------+----------+-----------+-----------+---------|
| MEAN 10574.5 4333.3     .00     .03|1.00   -.8| .97  -1.9|           | 72.3  70.6|         |
| S.D.   638.2   58.2     .68     .00| .13   4.4| .18   4.4|           |  4.0   2.4|         |
-------------------------------------------------------------------------------------------------
```

Empirical Fit Statistics: Middle School Instrument

Table 59.
Empirical Fit Statistics: Climate4Creativity Middle School Overall Measure.

TABLE 10.1 C4C/M Overall - All Schools ZOU248WS.TXT Apr 13 23:23 2015
INPUT: 3263 Student 126 Item REPORTED: 3221 Student 76 Item 5 CATS WINSTEPS 3.81.0

Student: REAL SEP.: 3.90 REL.: .94 ... Item: REAL SEP.: 25.30 REL.: 1.00
Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASURE-A CORR.	PTMEASURE-A EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
18	6528	3097	1.32	.02	1.47	9.9	1.62	9.9	A .31	.51	29.6	33.8	9C-STU
12	14259	3140	-1.22	.03	1.54	9.9	1.40	8.8	B .35	.38	59.2	60.2	8C-WKHD
61	13320	3076	-.82	.02	1.19	5.7	1.51	9.9	C .19	.41	41.4	48.0	23-FGRD
13	10661	3138	.20	.02	1.32	9.9	1.48	9.9	D .32	.50	28.8	33.7	8D-TTCH
16	11585	3146	-.05	.02	1.30	9.9	1.48	9.9	E .33	.48	33.0	35.7	9A-FRND
14	8458	3116	.77	.02	1.22	9.7	1.44	9.9	F .35	.52	30.7	30.8	8E-EVNT
80	11502	3077	-.10	.02	1.09	3.6	1.43	9.9	G .25	.48	39.3	36.3	29-FLRN
95	7930	2964	.81	.02	1.35	9.9	1.43	9.9	H .41	.53	26.3	30.9	34D-LTIM
71	13015	3029	-.76	.02	1.42	9.9	1.28	7.3	I .44	.41	43.2	46.5	26G-FINS
10	10095	3176	.39	.02	1.22	9.6	1.41	9.9	J .37	.51	30.5	32.6	8A-HWK
68	11985	3041	-.31	.02	1.28	9.7	1.21	6.7	K .44	.46	34.4	38.4	26E-FREC
81	10281	2999	.18	.02	1.13	5.3	1.28	9.9	L .30	.50	35.3	34.0	30-LGRD
20	9863	3111	.39	.02	1.19	8.3	1.26	9.9	M .45	.51	29.3	32.5	9E-ADLT
69	11475	3038	-.14	.02	1.26	9.6	1.24	8.2	N .44	.47	32.4	36.7	26H-FINV
65	13014	3067	-.69	.02	1.24	7.5	1.15	4.4	O .43	.42	42.5	44.7	26B-FENC
78	12091	2993	-.42	.02	1.23	7.7	1.18	5.6	P .44	.44	35.2	39.7	27G-FCRE
77	11789	3012	-.28	.02	1.17	6.2	1.19	6.2	Q .41	.46	34.0	38.1	27F-FTM
48	11390	3119	-.02	.02	1.18	7.0	1.17	6.2	R .49	.49	32.6	35.5	15E-TCH
55	11192	3030	-.07	.02	1.13	5.3	1.18	6.4	S .49	.48	32.1	35.7	20C-PUN
67	11979	3051	-.30	.02	1.18	6.4	1.10	3.2	T .48	.46	36.3	38.2	26D-FENH
70	12384	3036	-.47	.02	1.18	6.1	1.10	3.3	U .46	.44	38.7	40.5	26F-FTIM
74	12514	3031	-.53	.02	1.16	5.3	1.14	4.2	V .43	.44	39.4	41.4	27C-FDIF
56	11074	3025	-.03	.02	1.14	5.5	1.14	5.0	W .51	.48	32.0	35.5	20D-TALK
72	12422	3071	-.43	.02	1.11	3.8	1.14	4.4	X .38	.45	38.8	40.0	27A-FFUN
73	11978	3035	-.32	.02	1.13	4.7	1.12	3.8	Y .42	.46	38.7	38.5	27B-FTHK
82	12556	3017	-.57	.02	1.07	2.4	1.13	3.9	Z .32	.43	43.6	42.2	31-LEXP
76	12322	2999	-.51	.02	1.12	4.2	1.10	3.0	.44	.44	40.1	41.1	27E-FACT
91	10442	2951	.08	.02	1.12	5.1	1.06	2.4	.60	.50	29.7	34.6	33G-LINS
66	12351	3045	-.44	.02	1.11	4.0	1.05	1.5	.47	.45	39.9	40.0	26C-FRET
17	10014	3128	.37	.02	.96	-1.6	1.09	3.8	.39	.51	36.5	32.6	9B-CLS
59	9689	3033	.37	.02	1.08	3.5	1.09	3.6	.53	.51	30.3	32.6	21C-TELL
100	8935	3015	.57	.02	.80	-9.6	1.08	3.2	.41	.52	42.1	31.5	36-LLRN
46	10758	3117	.16	.02	1.06	2.8	1.04	1.6	.58	.50	32.9	34.1	15C-AFRD
BETTER FITTING OMITTED													
84	10716	3020	.08	.02	.95	-2.0	.91	-3.4	y .57	.50	35.5	34.7	33A-LDIR
87	9376	2988	.42	.02	.95	-2.1	.93	-2.9	x .62	.52	30.6	32.2	33D-LENH
94	9474	2968	.38	.02	.94	-2.7	.93	-3.0	w .59	.51	34.1	32.5	34C-LDIF
32	11948	3131	-.18	.02	.91	-3.8	.93	-2.7	v .46	.47	43.3	37.2	12F-AFTR
96	9215	2943	.43	.02	.93	-3.3	.91	-3.7	u .60	.52	32.7	32.2	34E-LACT
54	11995	3048	-.31	.02	.92	-3.1	.90	-3.4	t .55	.46	40.7	38.3	20B-HELP
86	9952	2980	.26	.02	.91	-4.1	.88	-5.0	s .62	.51	34.1	33.3	33C-LRET
35	11603	3135	-.07	.02	.88	-5.3	.89	-4.4	r .47	.48	42.4	35.9	13C-BATH
47	10371	3115	.27	.02	.88	-5.5	.87	-5.6	q .62	.51	36.2	33.4	15D-NICE
24	12009	3114	-.23	.02	.85	-6.2	.86	-5.2	p .55	.47	40.3	37.5	10D-RSTS
27	9780	3070	.38	.02	.79	-9.9	.85	-6.4	o .52	.51	41.2	32.6	11-LIKE
37	12115	3137	-.23	.02	.83	-7.3	.84	-5.9	n .48	.47	47.1	37.6	13E-STRS
28	12260	3167	-.24	.02	.78	-9.6	.82	-6.7	m .48	.47	46.7	37.7	12A-TRAV
39	12291	3132	-.30	.02	.80	-8.5	.82	-6.5	l .49	.46	47.9	38.2	13G-LRM
31	12126	3124	-.25	.02	.78	-9.2	.81	-7.3	k .49	.47	46.2	37.8	12D-LNCH
23	11669	3129	-.10	.02	.76	-9.9	.79	-8.4	j .55	.48	43.3	36.2	10C-KNTS
26	9460	3124	.51	.02	.73	-9.9	.79	-9.6	i .49	.52	42.2	31.7	10F-RSST
30	13105	3135	-.60	.02	.79	-8.0	.76	-7.9	h .53	.44	51.6	42.8	12C-CLS
25	9318	3106	.53	.02	.68	-9.9	.75	-9.9	g .50	.52	42.8	31.6	10E-KNST
34	11990	3148	-.18	.02	.72	-9.9	.74	-9.9	f .50	.47	47.4	37.1	13B-HALL
60	10817	3086	.11	.02	.71	-9.9	.73	-9.9	e .56	.50	44.9	34.4	22-SAFE
22	8823	3137	.69	.02	.66	-9.9	.72	-9.9	d .51	.52	42.8	31.1	10B-RSSS
33	13157	3166	-.57	.02	.71	-9.9	.72	-9.9	c .53	.44	54.1	42.2	13A-CLS
21	9223	3162	.61	.02	.66	-9.9	.71	-9.9	b .51	.52	43.1	31.5	10A-KNSS
29	12551	3136	-.39	.02	.62	-9.9	.65	-9.9	a .52	.46	53.7	39.3	12B-B4
MEAN	10981.4	3066.7	.00	.02	1.01	.1	1.04	.4			38.2	36.7	
S.D.	1559.0	64.3	.49	.00	.20	6.5	.21	6.2			7.5	6.1	

Table 60.
Empirical Fit Statistics: Climate4Creativity Middle School Safety Measure.

```
TABLE 10.1 C4C/M Safety - All Schools           ZOU806WS.TXT  Apr 13 23:26 2015
INPUT: 3263 Student  126 Item  REPORTED: 3205 Student  25 Item  5 CATS WINSTEPS 3.81.0
-------------------------------------------------------------------------------------------
Student: REAL SEP.: 2.91  REL.: .89 ... Item: REAL SEP.: 16.71  REL.: 1.00
         Item STATISTICS:  MISFIT ORDER
-------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL           MODEL|   INFIT  |  OUTFIT  |PTMEASURE-A|EXACT MATCH|         |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item    |
|------------------------------------+----------+----------+-----------+-----------+---------|
|    48  11390   3119     .04     .02|1.44   9.9|1.51   9.9|A .53   .60| 36.7  43.2| 15E-TCH |
|    59   9689   3033     .57     .02|1.33   9.9|1.43   9.9|B .56   .63| 32.3  38.8| 21C-TELL|
|    56  11074   3025     .03     .02|1.35   9.9|1.39   9.9|C .55   .59| 37.0  43.2| 20D-TALK|
|    55  11192   3030    -.02     .02|1.30   9.9|1.36   9.9|D .55   .59| 38.6  43.8| 20C-PUN |
|    58   9644   3050     .61     .02|1.25   9.9|1.34   9.9|E .57   .63| 32.8  38.4| 21B-STOP|
|    57  10025   3047     .46     .02|1.16   6.4|1.28   9.6|F .57   .63| 38.1  39.6| 21A-HELP|
|    53  10870   3073     .18     .02|1.04   1.7|1.17   5.7|G .60   .61| 42.9  41.9| 20A-STOP|
|    45  10530   3127     .38     .02|1.12   4.9|1.14   5.1|H .62   .62| 39.1  40.1| 15B-STFF|
|    46  10758   3117     .28     .02|1.14   5.4|1.12   4.2|I .64   .61| 39.5  40.9| 15C-AFRD|
|    38  10795   3126     .28     .02|1.03   1.4|1.10   3.6|J .58   .61| 42.2  40.9| 13F-PLOT|
|    54  11995   3048    -.34     .02|1.04   1.6|1.02    .6|K .59   .56| 46.8  47.1| 20B-HELP|
|    44  11232   3147     .14     .02| .98   -.8|1.00    .0|L .64   .60| 43.5  42.1| 15A-HURT|
|    28  12260   3167    -.25     .02| .87  -5.1| .98   -.8|M .57   .58| 52.8  46.2| 12A-TRAV|
|    32  11948   3131    -.17     .02| .95  -2.0| .95  -1.7|l .58   .58| 51.0  45.4| 12F-AFTR|
|    36  11248   3075     .04     .02| .90  -3.9| .93  -2.6|k .60   .60| 48.3  43.3| 13D-PLGR|
|    47  10371   3115     .43     .02| .91  -4.0| .93  -2.9|j .67   .62| 41.6  39.9| 15D-NICE|
|    30  13105   3135    -.74     .02| .91  -3.2| .90  -2.9|i .57   .54| 57.0  51.9| 12C-CLS |
|    35  11603   3135    -.02     .02| .87  -5.3| .91  -3.4|h .61   .59| 51.7  43.8| 13C-BATH|
|    31  12126   3124    -.26     .02| .80  -8.0| .82  -6.5|g .61   .57| 54.9  46.3| 12D-LNCH|
|    37  12115   3137    -.24     .02| .82  -7.2| .82  -6.5|f .61   .58| 55.4  46.2| 13E-STRS|
|    33  13157   3166    -.69     .02| .77  -8.4| .78  -7.3|e .60   .54| 59.9  51.6| 13A-CLS |
|    39  12291   3132    -.33     .02| .77  -9.1| .76  -8.6|d .62   .57| 58.0  47.0| 13G-LRM |
|    60  10817   3086     .22     .02| .69  -9.9| .76  -9.6|c .66   .61| 53.7  41.5| 22-SAFE |
|    34  11990   3148    -.16     .02| .69  -9.9| .75  -9.4|b .63   .58| 56.0  45.3| 13B-HALL|
|    29  12551   3136    -.44     .02| .64  -9.9| .69  -9.9|a .62   .56| 60.0  48.7| 12B-B4  |
|------------------------------------+----------+----------+-----------+-----------+---------|
| MEAN 11391.0 3105.2     .00     .02| .99   -.6|1.03    .3|           | 46.8  43.9|         |
| S.D.   943.4   43.5     .36     .00| .22   7.1| .23   6.9|           |  8.6   3.6|         |
-------------------------------------------------------------------------------------------
```

Table 61.
Empirical Fit Statistics: Climate4Creativity Middle School Community Measure.

```
TABLE 10.1 C4C/M Community - All Schools         ZOU653WS.TXT  Apr 13 23:27 2015
INPUT: 3263 Student  126 Item  REPORTED: 3218 Student  16 Item  5 CATS WINSTEPS 3.81.0
-------------------------------------------------------------------------------------------
Student: REAL SEP.: 2.12  REL.: .82 ... Item: REAL SEP.: 32.51  REL.: 1.00
         Item STATISTICS:  MISFIT ORDER
-------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL           MODEL|   INFIT  |  OUTFIT  |PTMEASURE-A|EXACT MATCH|         |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item    |
|------------------------------------+----------+----------+-----------+-----------+---------|
|    12  14259   3140   -1.72     .03|1.45   9.9|1.30   6.7|A .43   .45| 62.8  62.2| 8C-WKHD |
|    18   6528   3097    1.31     .02|1.40   9.9|1.38   9.9|B .49   .54| 33.3  38.7| 9C-STU  |
|    13  10661   3138    -.07     .02|1.34   9.9|1.37   9.9|C .46   .55| 30.7  37.8| 8D-TTCH |
|    16  11585   3146    -.39     .02|1.29   9.9|1.28   9.9|D .49   .54| 36.5  38.3| 9A-FRND |
|    14   8458   3116     .64     .02|1.22   8.8|1.27   9.9|E .50   .56| 34.1  36.9| 8E-EVNT |
|    10  10095   3176     .16     .02|1.21   8.6|1.23   9.0|F .53   .56| 34.3  38.0| 8A-HWK  |
|    20   9863   3111     .16     .02|1.21   8.3|1.20   7.7|G .58   .56| 32.9  38.0| 9E-ADLT |
|    27   9780   3070     .15     .02| .98   -.9|1.09   3.5|H .52   .56| 42.9  37.9| 11-LIKE |
|    19  12837   3125    -.93     .02|1.02    .9| .98   -.7|h .57   .51| 46.0  42.5| 9D-TCHR |
|    24  12009   3114    -.60     .02| .96  -1.7| .96  -1.4|g .58   .53| 41.5  39.2| 10D-RSTS|
|    17  10014   3128     .13     .02| .91  -4.0| .93  -3.0|f .56   .56| 42.4  37.6| 9B-CLS  |
|    23  11669   3129    -.45     .02| .84  -6.8| .84  -6.8|e .60   .54| 44.6  38.3| 10C-KNTS|
|    26   9460   3124     .31     .02| .73  -9.9| .73  -9.9|d .60   .56| 45.9  37.6| 10F-RSST|
|    25   9318   3106     .34     .02| .67  -9.9| .68  -9.9|c .61   .56| 46.9  37.4| 10E-KNST|
|    21   9223   3162     .43     .02| .64  -9.9| .66  -9.9|b .62   .56| 48.6  37.3| 10A-KNSS|
|    22   8823   3137     .54     .02| .64  -9.9| .66  -9.9|a .62   .56| 50.0  37.3| 10B-RSSS|
|------------------------------------+----------+----------+-----------+-----------+---------|
| MEAN 10286.4 3126.2     .00     .02|1.03    .8|1.04    .9|           | 42.1  39.7|         |
| S.D.  1801.1   24.4     .68     .00| .27   8.1| .25   8.0|           |  8.1   5.9|         |
-------------------------------------------------------------------------------------------
```

Table 62.
Empirical Fit Statistics: Climate4Creativity Middle School Higher Order Learning Measure.

TABLE 10.1 C4C/M Learning & Creativity - All Sch ZOU458WS.TXT Apr 13 23:29 2015
INPUT: 3263 Student 126 Item REPORTED: 3132 Student 34 Item 3 CATS WINSTEPS 3.81.0

Student: REAL SEP.: 3.10 REL.: .91 ... Item: REAL SEP.: 23.48 REL.: 1.00
Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASURE-A CORR.	PTMEASURE-A EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
62	8196	3086	-1.83	.04	1.21	8.3	1.57	9.6	A .32	.48	67.5	72.8	24-FEXP
80	6794	3077	.18	.04	1.18	6.9	1.48	9.9	B .27	.57	66.3	69.2	29-FLRN
82	7236	3017	-.62	.04	1.24	9.5	1.40	9.9	C .34	.54	60.6	67.9	31-LEXP
95	5500	2964	1.60	.04	1.28	9.9	1.30	9.6	D .47	.57	66.9	67.3	34D-LTIM
81	6325	2999	.59	.04	1.10	3.7	1.27	8.8	E .30	.57	70.7	69.7	30-LGRD
72	7344	3071	-.59	.04	1.08	3.5	1.16	4.8	F .45	.54	65.0	68.0	27A-FFUN
100	5950	3015	1.14	.04	.79	-8.9	1.12	4.1	G .37	.57	78.0	69.2	36-LLRN
69	7044	3038	-.29	.04	1.08	3.1	1.08	2.6	H .54	.55	67.3	68.3	26H-FINV
71	7814	3029	-1.45	.04	1.08	3.6	1.07	1.6	I .54	.51	73.3	70.1	26G-FINS
68	7275	3041	-.61	.04	1.07	3.0	1.05	1.6	J .54	.54	67.3	68.0	26E-FREC
77	7120	3012	-.48	.04	1.07	2.8	1.07	2.3	K .49	.55	68.4	68.1	27F-FTM
70	7394	3036	-.79	.04	1.06	2.6	1.05	1.6	L .53	.54	68.6	68.2	26F-FTIM
91	6537	2951	.14	.04	1.06	2.1	1.01	.4	M .66	.56	67.7	69.2	33G-LINS
64	8064	3088	-1.61	.04	1.00	.0	1.05	1.2	N .47	.50	71.6	71.0	26A-FDIR
65	7742	3067	-1.19	.04	1.02	.8	1.03	.8	O .53	.52	70.9	69.0	26B-FENC
67	7255	3051	-.54	.04	1.02	.7	1.00	-.1	P .56	.54	69.7	68.0	26D-FENH
73	7167	3035	-.46	.04	1.01	.4	1.02	.8	Q .51	.55	69.6	68.1	27B-FTHK
74	7455	3031	-.89	.04	1.01	.4	1.00	-.1	q .53	.53	70.6	68.3	27C-FDIF
78	7302	2993	-.80	.04	1.01	.5	.99	-.3	p .55	.53	70.4	68.1	27G-FCRE
66	7364	3045	-.71	.04	.97	-1.3	1.00	.1	o .55	.54	70.2	68.0	26C-FRET
90	6433	2984	.39	.04	.99	-.3	.96	-1.2	n .63	.57	70.1	69.7	33F-LTIM
76	7356	2999	-.86	.04	.97	-1.4	.98	-.6	m .53	.53	70.6	68.2	27E-FACT
98	5982	2947	.90	.04	.94	-2.4	.92	-3.0	l .67	.57	70.6	69.7	34G-LCRE
92	5693	3003	1.45	.04	.93	-2.7	.93	-2.5	k .61	.57	71.1	68.0	34A-LFUN
97	5969	2958	.95	.04	.91	-3.6	.89	-4.0	j .62	.57	72.0	69.6	34F-LTM
84	6633	3020	.22	.04	.89	-4.3	.87	-4.7	i .62	.56	72.4	69.4	33A-LDIR
88	6063	2974	.87	.04	.89	-4.3	.87	-4.8	h .69	.57	70.0	69.7	33E-LREC
93	5969	2970	.98	.04	.89	-4.2	.87	-4.6	g .64	.57	72.7	69.6	34B-LTHK
89	6099	2976	.82	.04	.88	-4.7	.85	-5.3	f .68	.57	72.2	69.8	33H-LINV
85	6366	2992	.50	.04	.87	-5.1	.84	-5.8	e .68	.57	72.5	69.8	33B-LENC
87	6133	2988	.81	.04	.86	-5.6	.83	-6.1	d .68	.57	72.4	69.8	33D-LENH
94	6113	2968	.78	.04	.85	-5.9	.83	-6.2	c .66	.57	73.3	69.8	34C-LDIF
86	6316	2980	.54	.04	.82	-7.4	.79	-7.7	b .68	.57	74.0	69.8	33C-LRET
96	5996	2943	.87	.04	.82	-7.4	.79	-7.6	a .68	.57	74.0	69.7	34E-LACT
MEAN	6764.7	3010.2	.00	.04	1.00	-.2	1.03	.1			70.3	69.1	
S.D.	716.7	40.5	.91	.00	.12	4.7	.19	5.1			3.1	1.1	

Appendix I – Empirical Item-person Maps for All Measures

Empirical item-person maps are shown in this appendix for each construct in both instruments. This appendix is included for the purposes of providing complete technical documentation on the instruments developed and tested in this study. Note that Winsteps *table 1.4* Wright maps are provided (Linacre, 2012) in order to demonstrate the range for polytomous scales. These maps show the distribution of persons versus difficulty of selecting items based on polytomous response, using the top 50%, middle 50% and bottom 50% of overall respondent scores. This provides both a typical graphical display of item difficulty versus person ability, and a clear perspective on the effective range of the measure.

Empirical Item-person Maps: Elementary Instrument

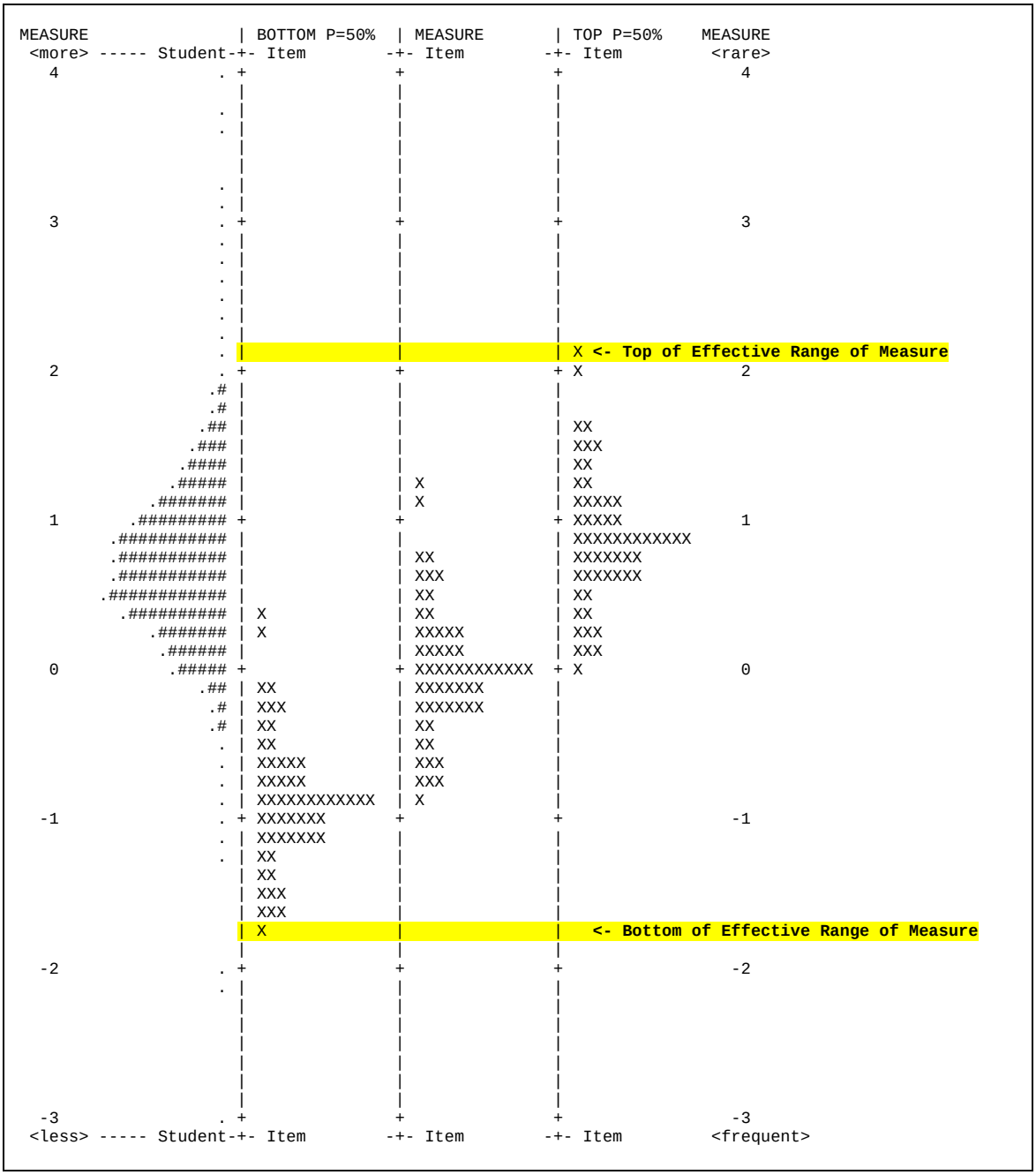


Figure 59. Empirical Item-person Map: Elementary Overall.

Note. "#" represents 38 students. "." Represents 1-37 students. "X" represents 1 item.

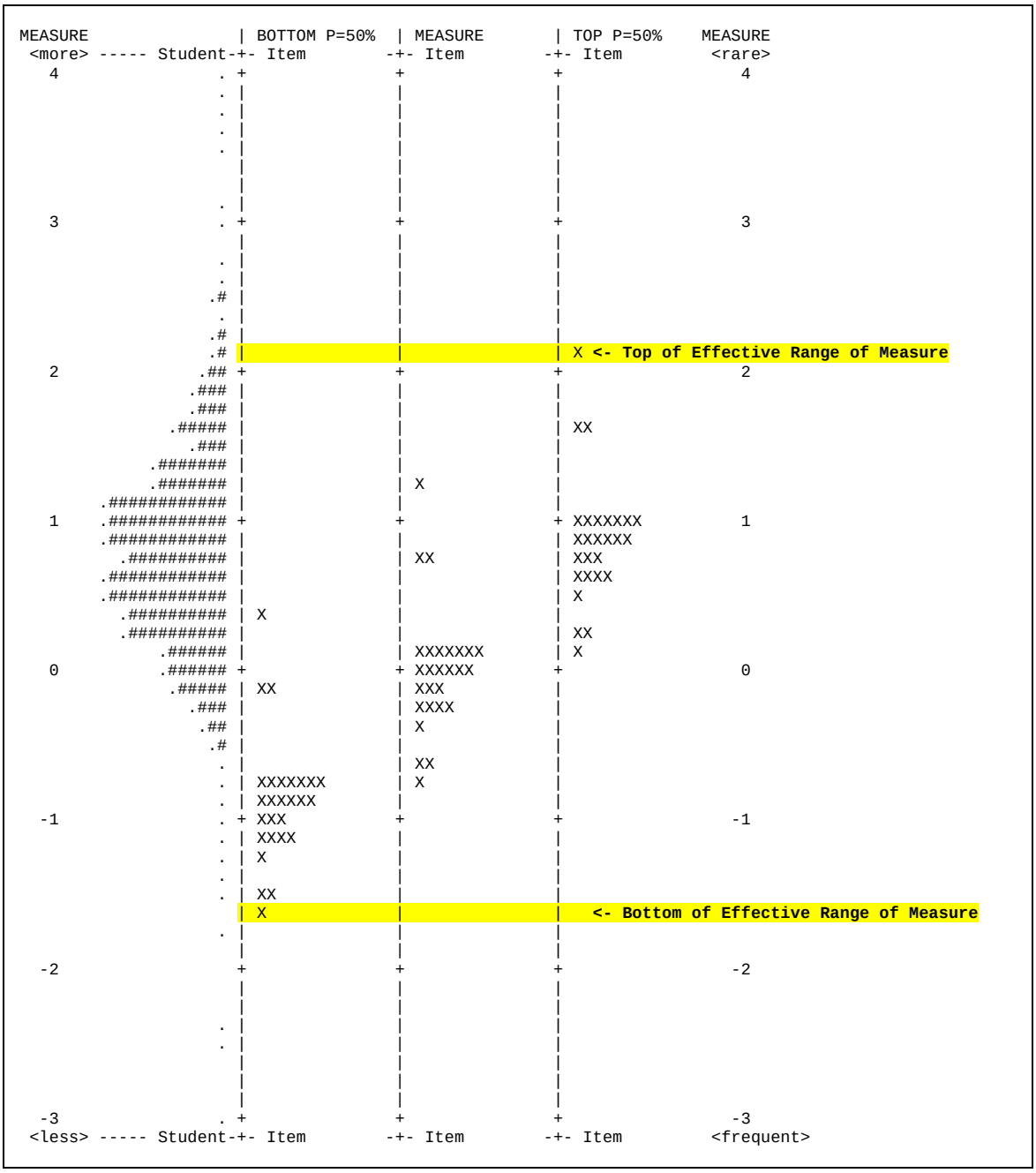


Figure 60. Empirical Item-person Map: Elementary Safety.

Note. "#" represents 28 students. "." Represents 1-27 students.

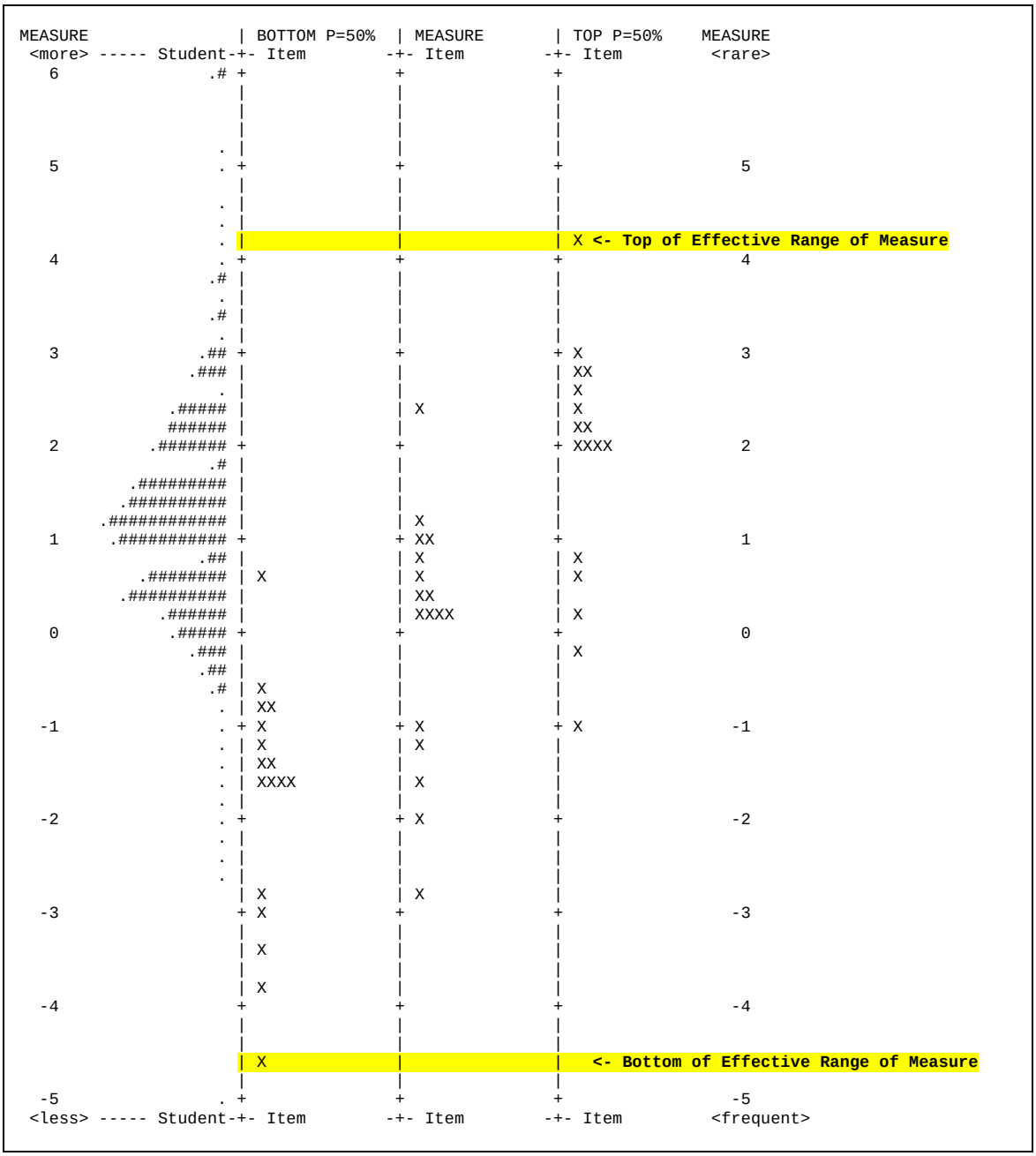


Figure 61. Empirical Item-person Map: Elementary Community.

Note. "#" represents 38 students. "." Represents 1-37 students.

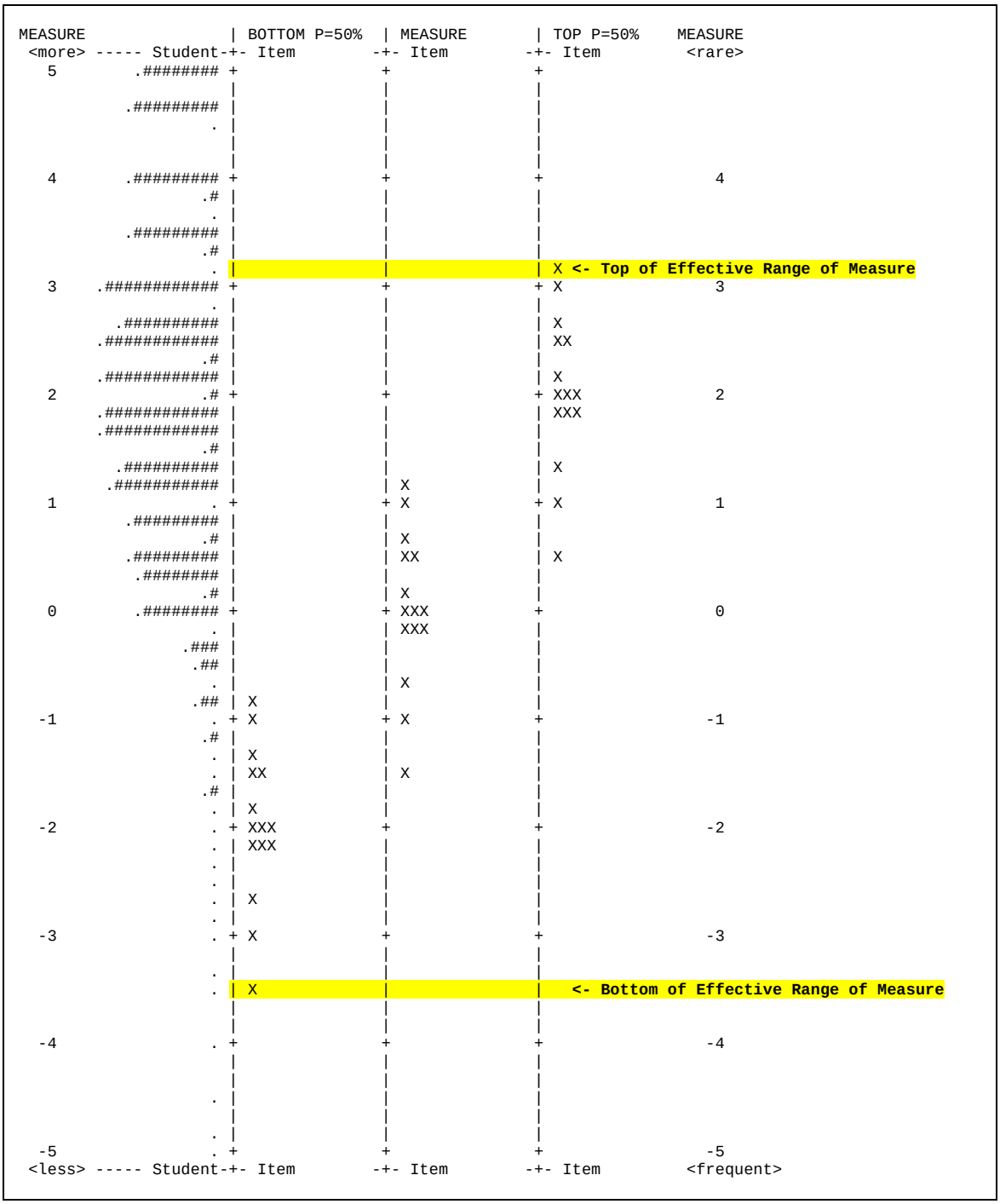


Figure 62. Empirical Item-person Map: Elementary H/O Learning.

Note. "#" represents 23 students. "." Represents 1-22 students.

Empirical Item-person Maps: Middle School Instrument

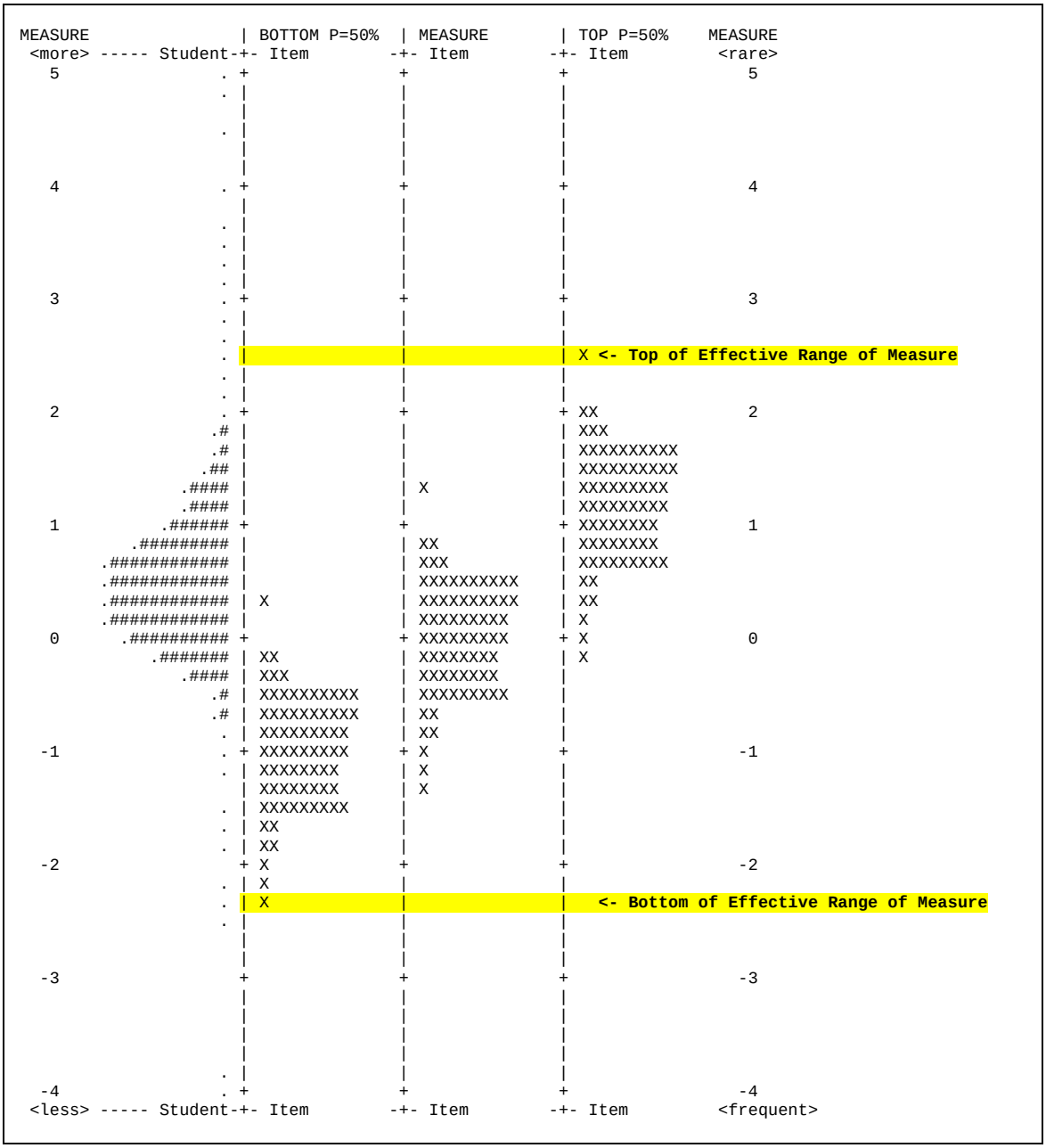


Figure 63. Empirical Item-person Map: Middle School Overall.

Note. "#" represents 29 students. "." Represents 1-28 students.

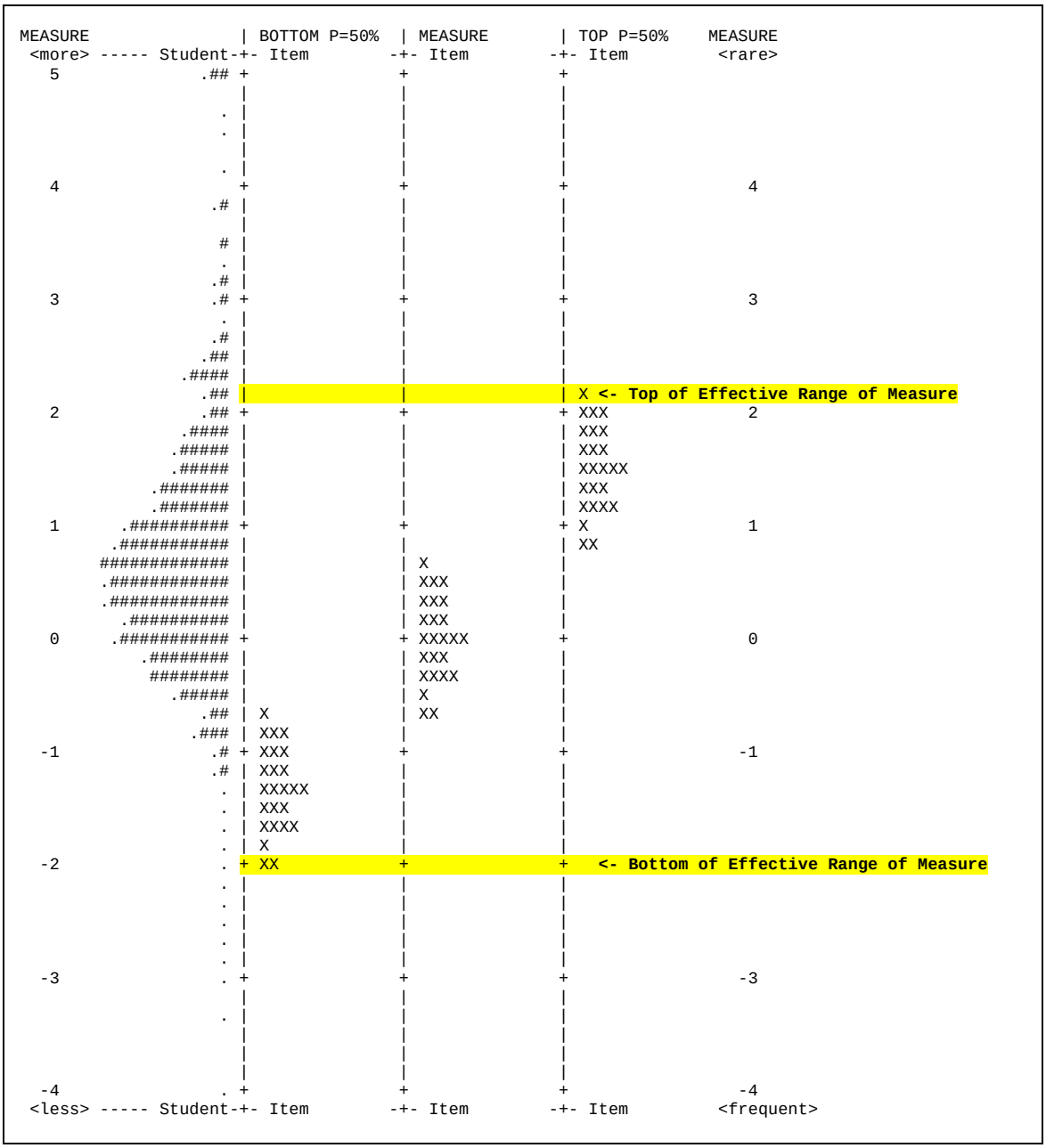


Figure 64. Empirical Item-person Map: Middle School Safety.

Note. "#" represents 19 students. "." Represents 1-18 students.

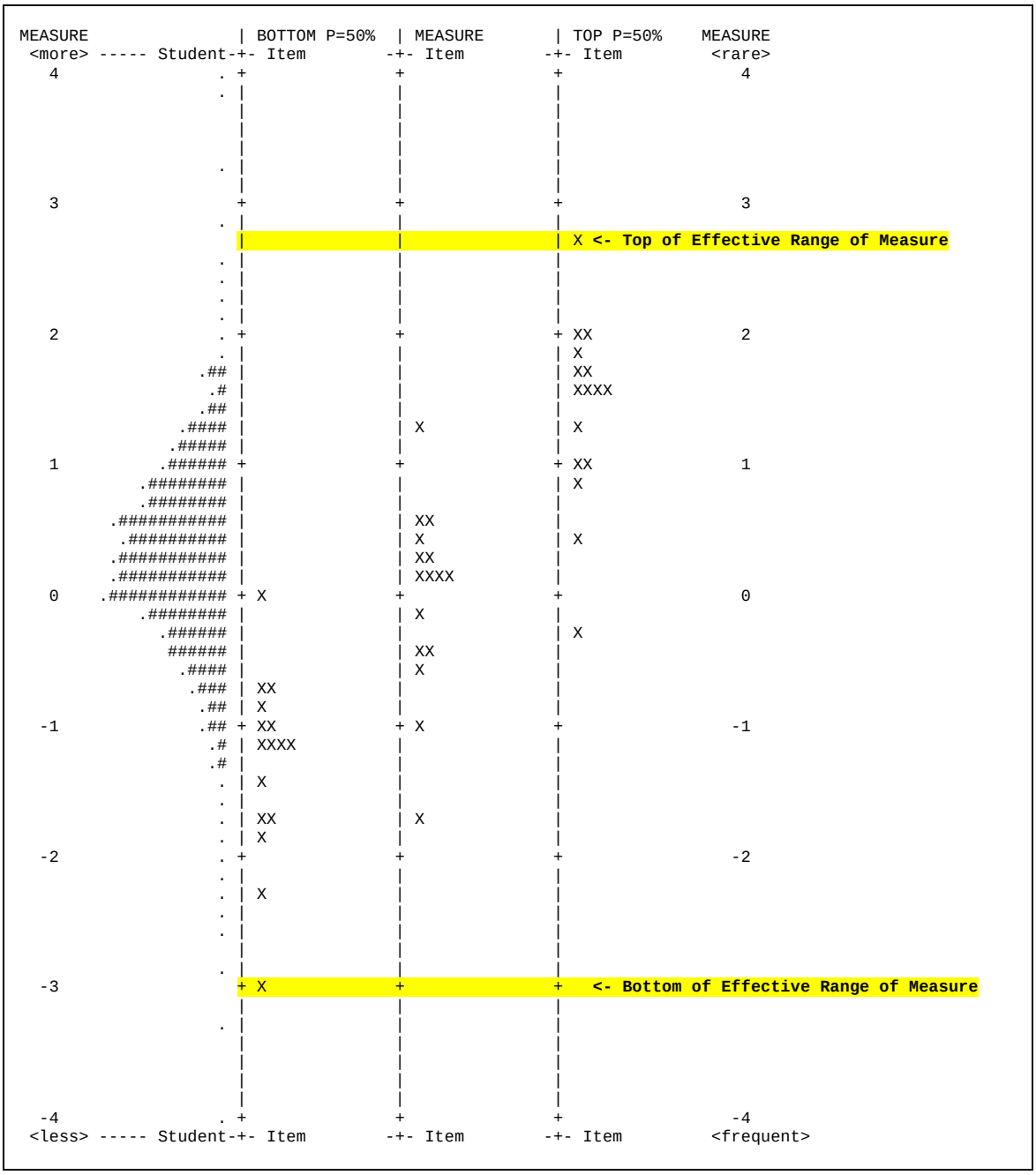


Figure 65. Empirical Item-person Map: Middle School Community.

Note. "#" represents 23 students. "." Represents 1-22 students.

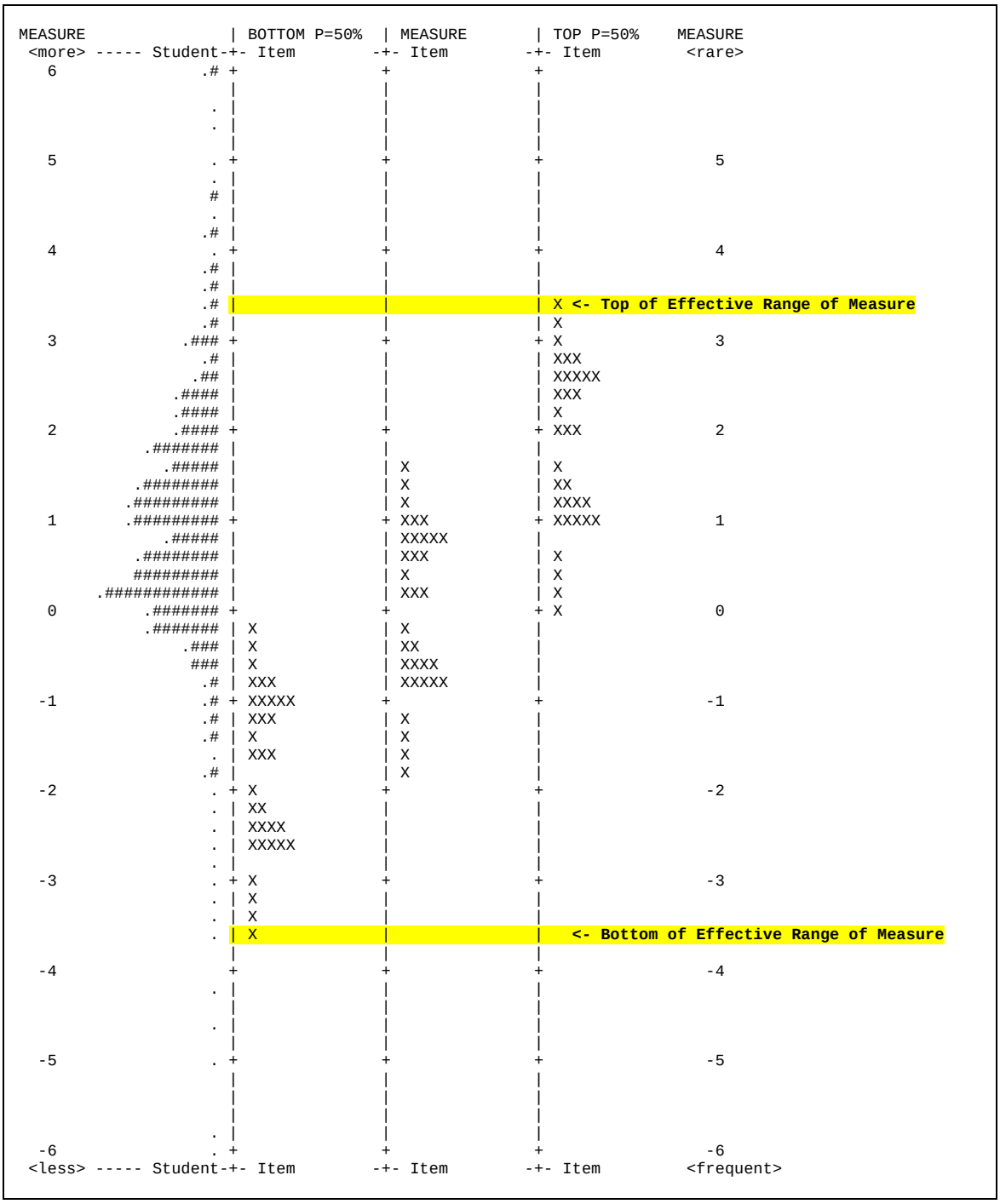


Figure 66. Empirical Item-person Map: Middle School H/O Learning.

Note. "#" represents 22 students. "." Represents 1-21 students.

Appendix J – Empirical Standard Error of Measurement Curves for All Measures

Empirical standard error of measurement curves are shown in this appendix for each construct in both instruments. This appendix is included for the purposes of providing complete technical documentation on the instruments developed and tested in this study. Note that Winsteps standard error data, plotted using Microsoft Excel (Microsoft, 2013a) are provided (Linacre, 2012) in order to document the standard errors for each measure.

Empirical SEM Curves: Elementary Instrument

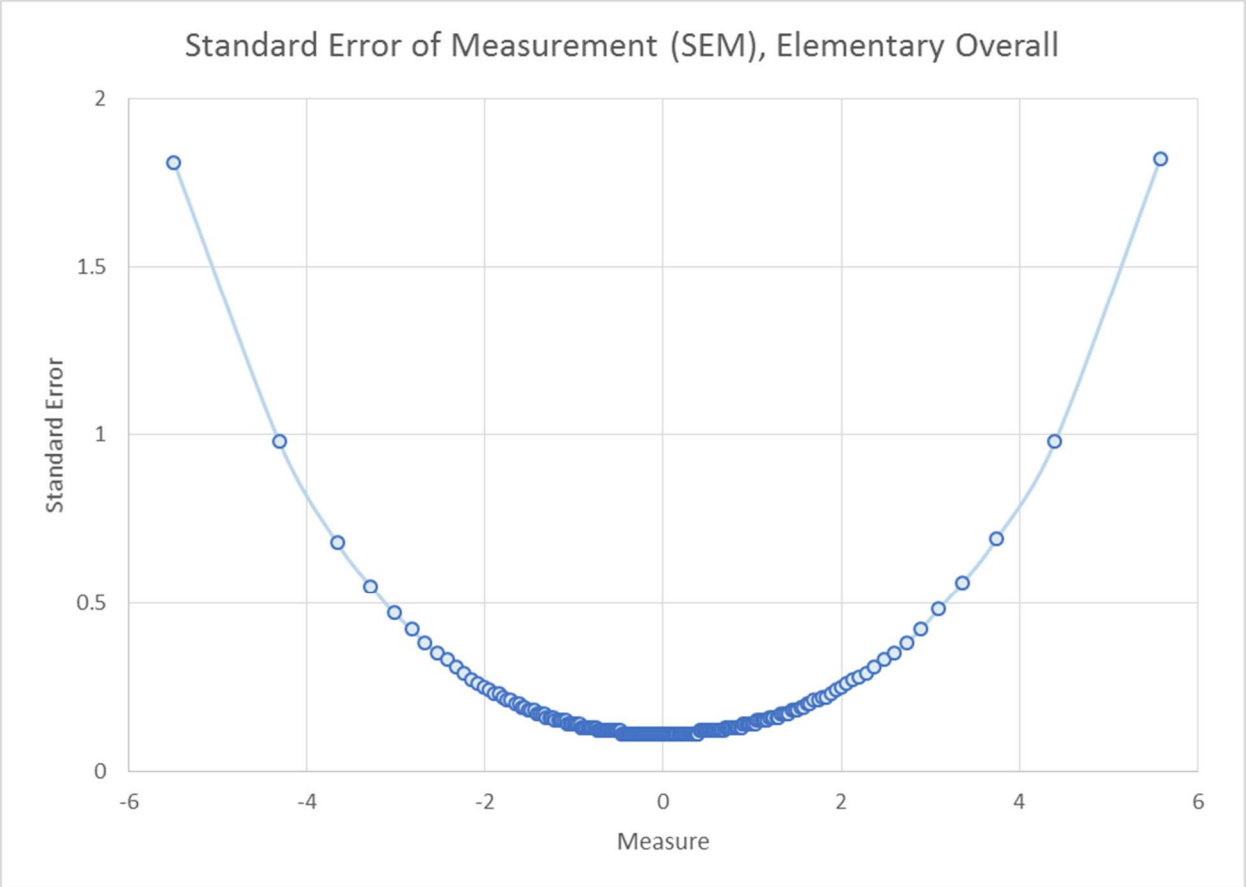


Figure 67. Empirical SEM Curve: Elementary Overall Measure.

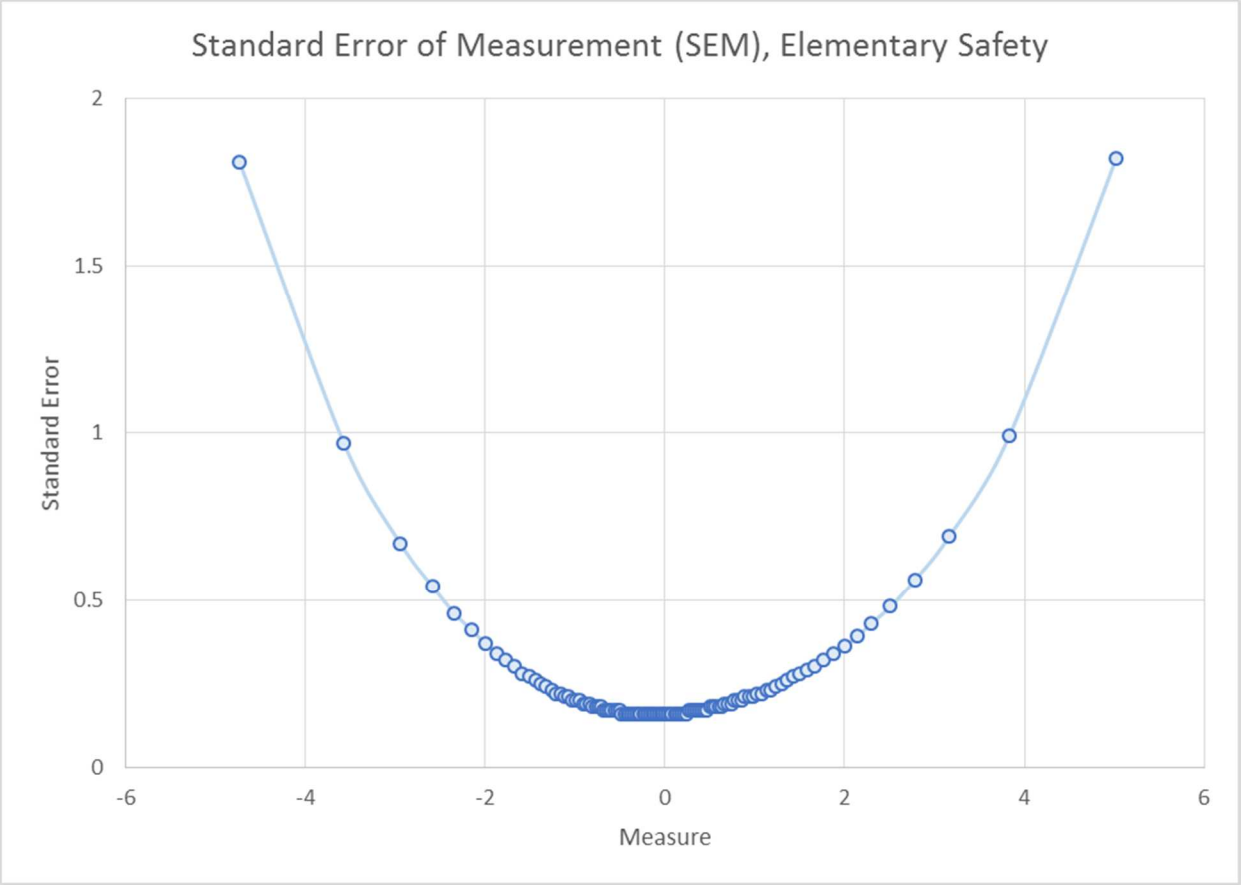


Figure 68. Empirical SEM Curve: Elementary Safety Measure.

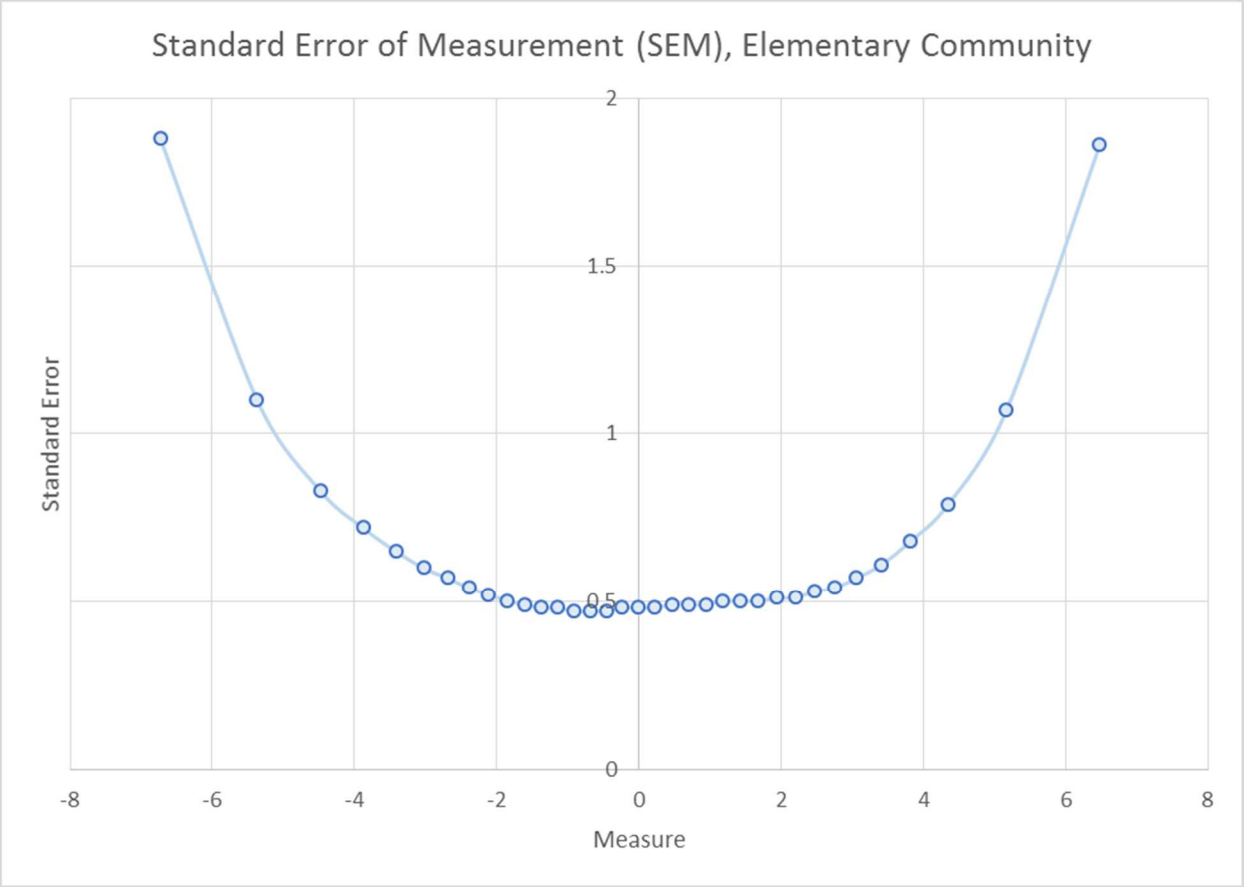


Figure 69. Empirical SEM Curve: Elementary Community Measure.

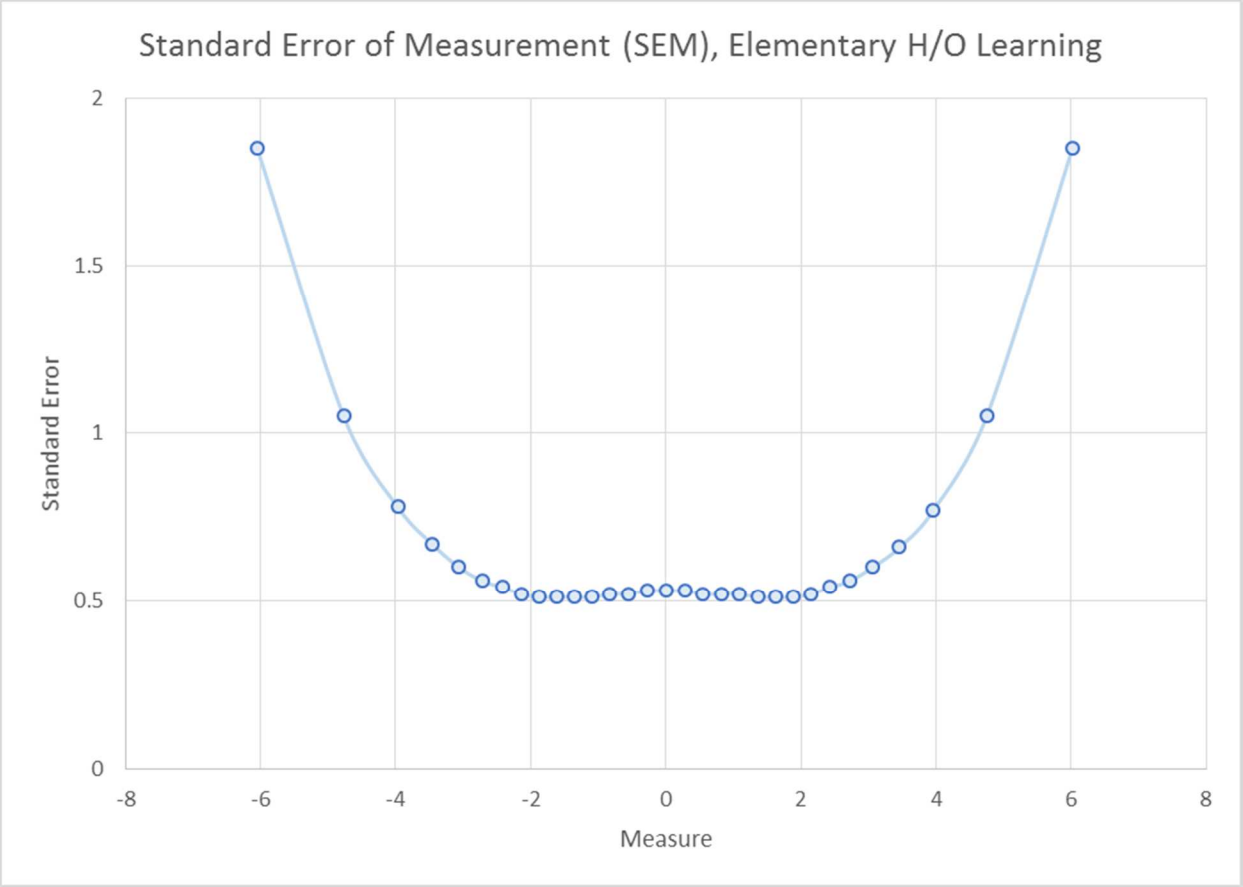


Figure 70. Empirical SEM Curve: Elementary Higher Order Learning Measure.

Empirical SEM Curves: Middle School Instrument

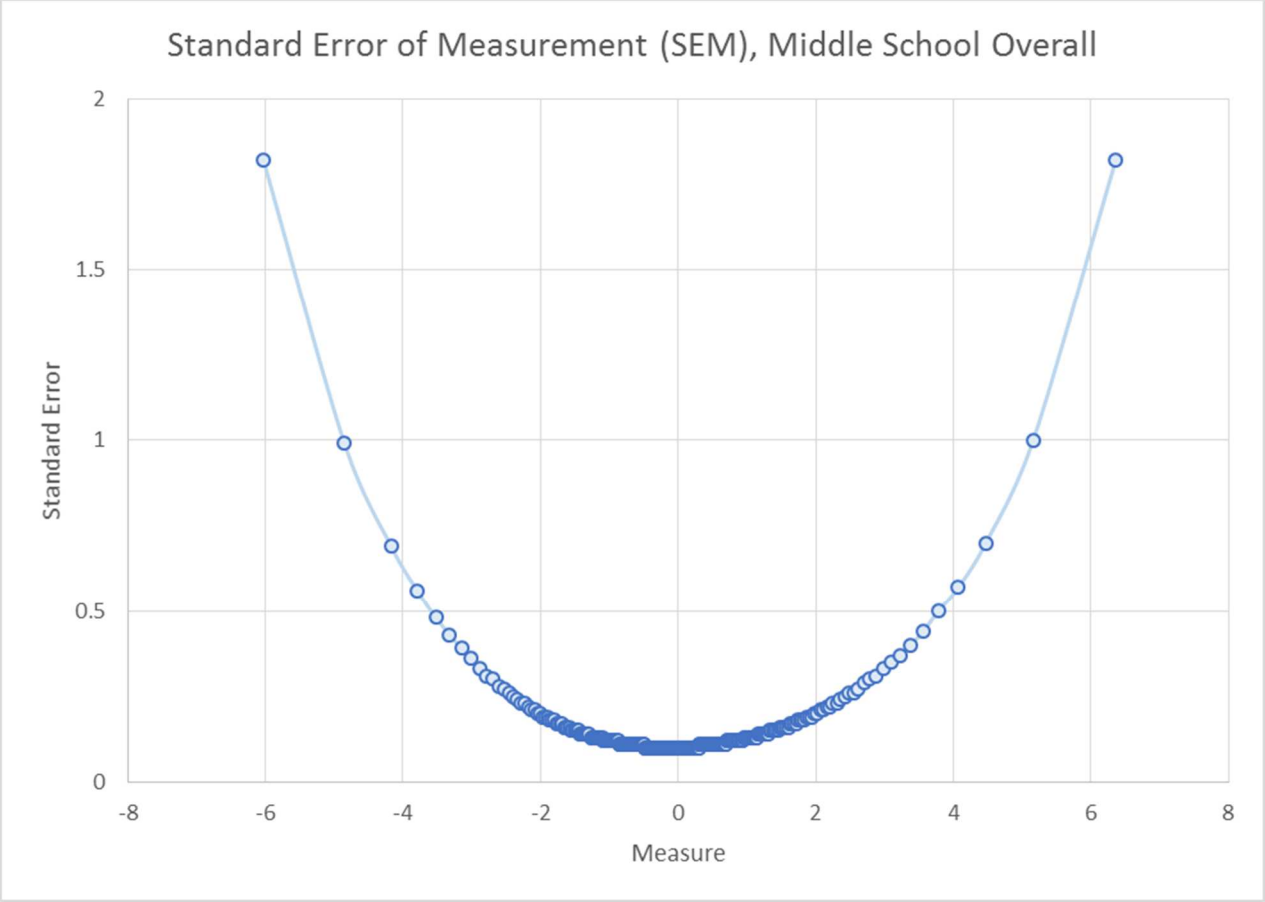


Figure 71. Empirical SEM Curve: Middle School Overall Measure.

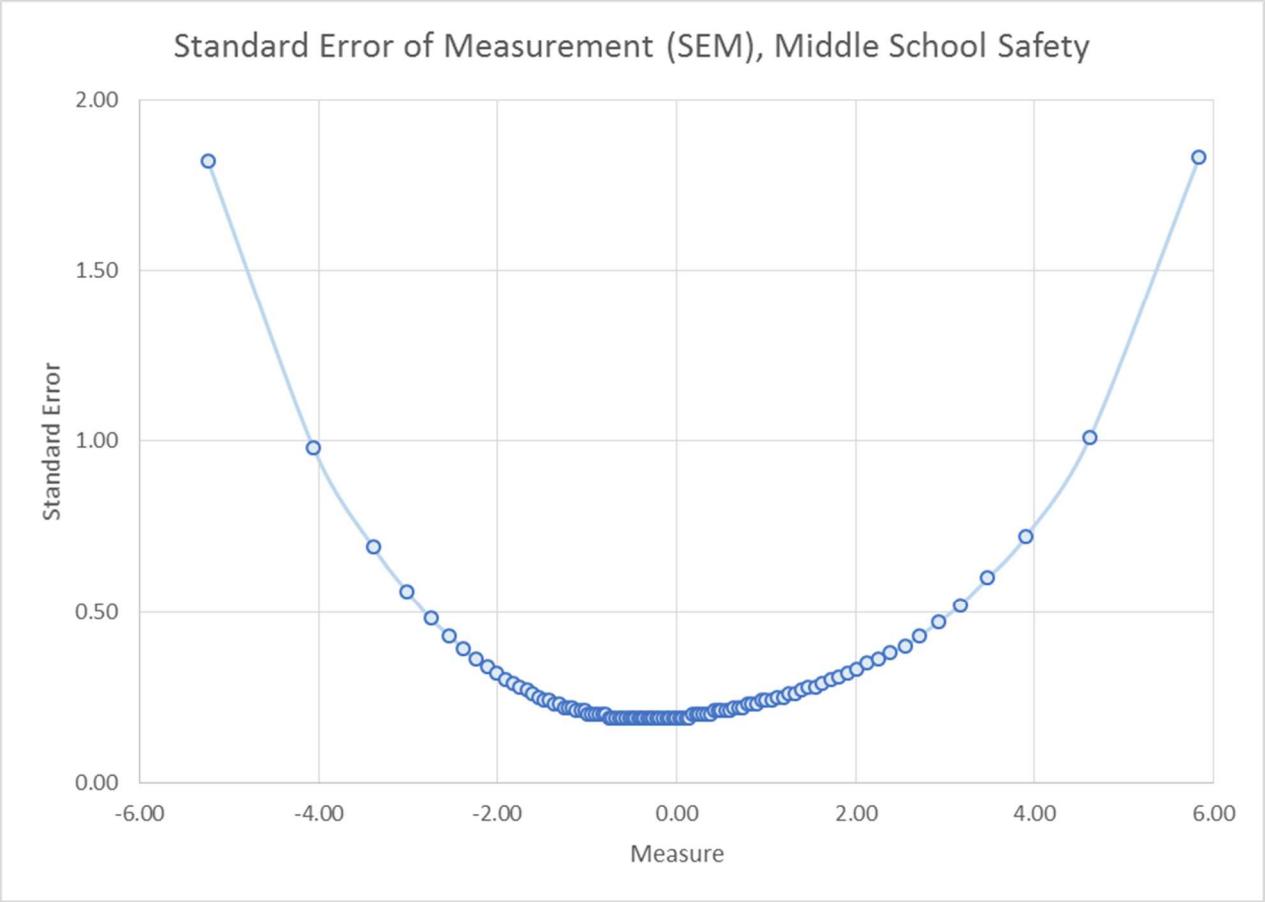


Figure 72. Empirical SEM Curve: Middle School Safety Measure.

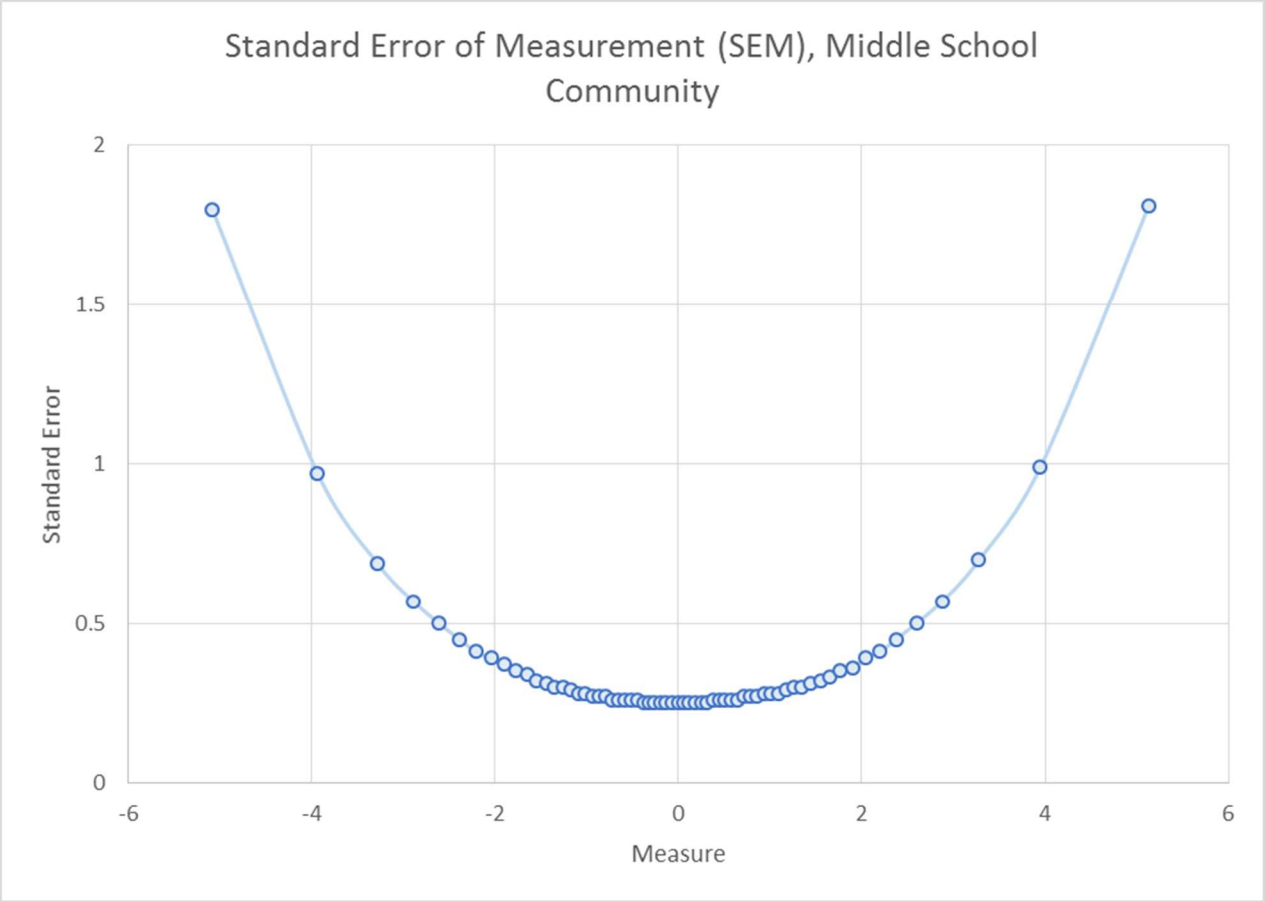


Figure 73. Empirical SEM Curve: Middle School Community Measure.

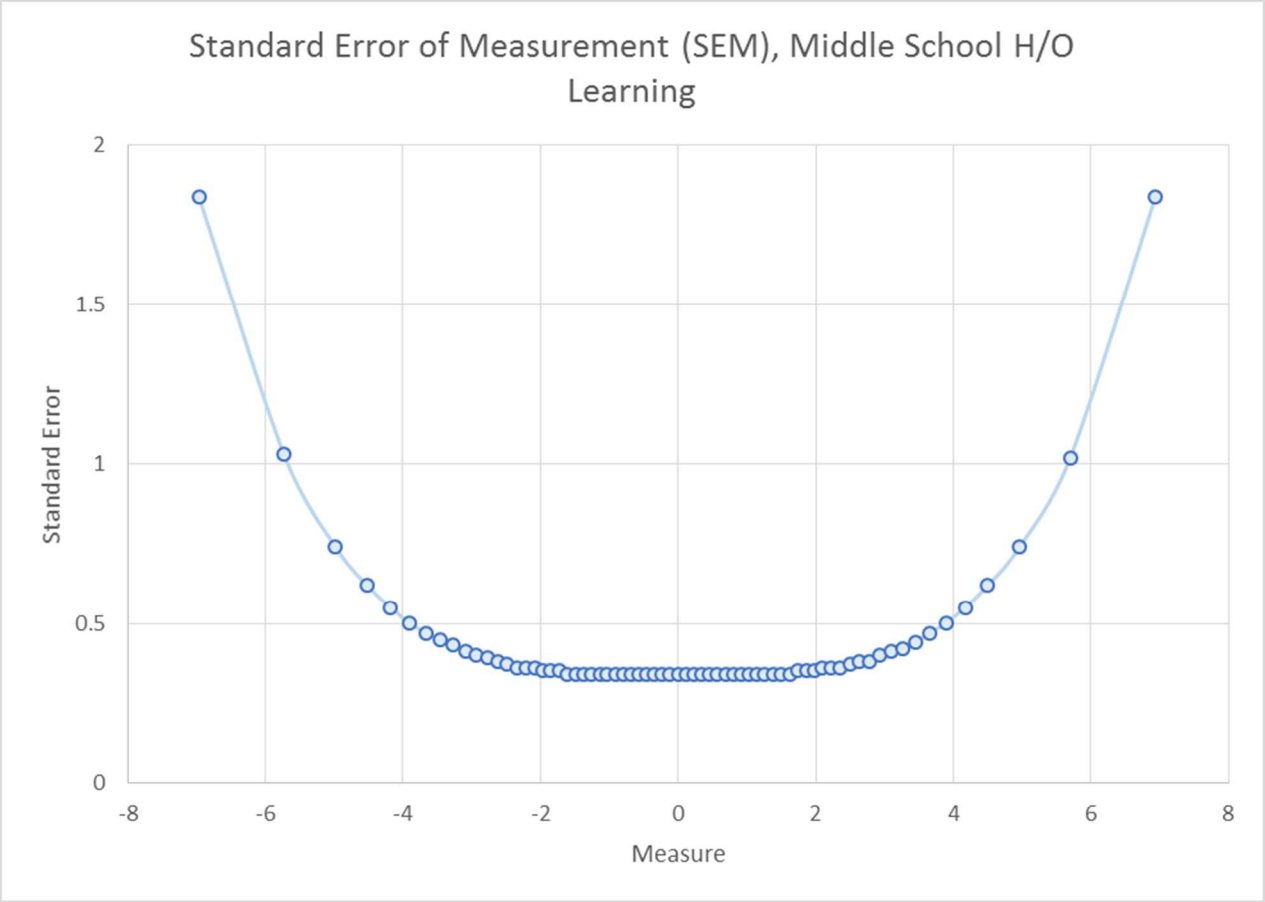


Figure 74. Empirical SEM Curve: Middle School Higher Order Learning Measure.

Appendix K – Professional Standards for Evaluation of Validity

Overarching Standard 1.0

"Clear articulation of each intended test score interpretation for a specified use should be set forth, and appropriate validity evidence in support of each intended interpretations should be provided." (American Educational Research Association et al., 2014, p. 23).

Cluster 1 – Establishing Intended Uses and Interpretations

Table 63.
Standards for Cluster 1 Establishing Intended Uses and Interpretations.

Standard	Ref.[1]	Standard Text
1.1	p. 23	The test developer should set forth clearly how test scores are intended to be interpreted and consequently used. The populations(s) for which a test is intended should be delimited clearly, and the construct or constructs that the test is intended to assess should be described clearly.
1.2	p. 23	A rationale should be presented for each individual interpretation of test scores for a given use, together with a summary of the evidence and theory bearing on the intended interpretation.
1.3	p. 23-24	If validity for some common or likely interpretation for a given use has not be evaluated, or if such an interpretation is inconsistent with available evidence, that fact should be made clear and potential users should be strongly cautioned about making unsupported interpretations.
1.4	p. 24	If a test score is interpreted for a given use in a way that has not been validated, it is incumbent on the user to justify the new interpretation for that use, providing a rationale and collecting new evidence, if necessary.
1.5	p. 24	When it is clearly stated or implied that a recommended test score interpretation for a given use will result in a specific outcome, the basis for expecting that outcome should be presented, together with relevant evidence.
1.6	p. 24	When a test use is recommended on the grounds that testing or the testing program itself will have some indirect benefit, in addition to the utility of information from interpretations of the test scores themselves, the recommender should make explicit the rationale for anticipating the indirect benefit. Logical or theoretical arguments and empirical evidence for the indirect benefit should be provided. Appropriate weight should be given to any contradictory findings in the scientific literature, including findings suggesting important indirect outcomes other than those predicted.
1.7	p. 24-25	If test performance, or a decision made therefrom, is claimed to be essentially unaffected by practice and coaching, then the propensity for test performance to change with these forms of instruction should be documented.

Note. [1]In American Educational Research Association et al. (2014).

Cluster 2 – Issues Regarding Samples and Settings Used in Validation

Table 64.
Standards for Cluster 2 Samples and Settings Used in Validation.

Standard	Ref.[1]	Standard Text
1.8	p. 25	The composition of any sample of test takers from which validity evidence is obtained should be described in as much detail as is practical and permissible, including major relevant socio-demographic and developmental characteristics.
1.9	p. 25-26	When a validation rests in part on the opinions or decisions of expert judges, observers, or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described. The qualifications and experience of the judges should be presented. The description of procedures should include any training and instructions provided, should indicate whether participants reached their decisions independently, and should report the level of agreement reached. If participants interacted with one another or exchanged information, the procedures through which they may have influenced one another should be set forth.
1.10	p. 26	When validity evidence includes statistical analyses of test results, either alone or together with the data on other variables, the conditions under which the data were collected should be described in enough detail that users can judge the relevance of the statistical findings to local conditions. Attention should be drawn to any features of a validation data collection that are likely to differ from typical operational testing conditions and that could plausibly influence test performance.

Note. [1]In American Educational Research Association et al. (2014).

Cluster 3 – Specific Forms of Validity Evidence

Table 65.
Standards for Cluster 3(a) Content-oriented Evidence.

Standard	Ref.[1]	Standard Text
1.11	p. 26	When the rationale for a test score interpretation for a given use rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporated criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified.

Note. [1]In American Educational Research Association et al. (2014).

Table 66.
Standards for Cluster 3(b) Cognitive Processes.

Standard	Ref.[1]	Standard Text
1.12	p. 26	If the rationale for score interpretation for a given use depends on premises about the psychological processes or cognitive operations of test takers, then theoretical or empirical evidence in support of those premises should be provided. When statements about the processes employed by observers and scorers are part of the argument for validity, similar information should be provided.

Note. [1]In American Educational Research Association et al. (2014).

Table 67.
Standards for Cluster 3(c) Internal Structure.

Standard	Ref.[1]	Standard Text
1.13	p. 26-27	If the rationale for a test score interpretation for a given use depends on premises about the relationships among test items or among parts of the test, evidence concerning the internal structure of the test should be provided.
1.14	p. 27	When interpretation of subscores, score differences, or profiles is suggested, the rationale and relevant evidence in support of such interpretation should be provided. When composite scores are developed, the basis for arriving at the composites should be given.
1.15	p.27	When interpretation of performance on specific items, or small subsets of items, is suggested, the rationale and relevant evidence in support of such interpretation should be provided. When interpretation of individual item responses is likely but is not recommended by the developer, the user should be warned against making such interpretations.

Note. [1]In American Educational Research Association et al. (2014).

Table 68.
Standards for Cluster 3(d) Relationships with Related Constructs

Standard	Ref.[1]	Standard Text
1.16	p. 27-28	When validity evidence includes empirical analysis of responses to test items together with data on other variables, the rationale for selecting the additional variables should be provided. Where appropriate and feasible, evidence concerning the constructs represented by other variables, as well as their technical properties, should be presented or cited. Attention should be drawn to any likely sources of dependence (or lack of independence) among variables other than dependencies among the construct(s) they represent.

Note. [1]In American Educational Research Association et al. (2014).

Table 69.
Standards for Cluster 3(e) Relationships with Criteria.

Standard	Ref.[1]	Standard Text
1.17	p. 28	When validation relies on evidence that test scores are related to one or more criterion variables, information about the suitability and technical quality of the criteria should be reported.
1.18	p. 28	When it is asserted that a certain level of test performance predicts adequate or inadequate criterion performance, information about the levels of criterion performance associated with given levels of test scores should be provided.
1.19	p. 28	If test scores are used in conjunction with other variables to predict some outcome or criterion, analyses based on statistical models of the predictor-criterion relationship should include those additional relevant variables along with the test scores.
1.20	p. 29	When effect size measures (e.g., correlations between test scores and criterion measures, standardized mean score differences between subgroups) are used to draw inferences that go beyond describing the sample or samples on which data have been collected, indices of the degree of uncertainty associated with these measures (e.g., standard errors, confidence intervals, or significance tests) should be reported.
1.21	p. 29	When statistical adjustments, such as those for restriction of a range or attenuation, are made, both adjusted and unadjusted coefficients, as well as the specific procedure used, and all statistics used in the adjustment, should be reported. Estimates of the construct-criterion relationship that remove the effects of measurement error on the test should be clearly reported as adjusted estimates.
1.22	p. 29-30	When a metaanalysis is used as evidence of the strength of a test-criterion relationship, the test and the criterion variables in the local situation should be comparable with those in the studies summarized. If relevant research includes credible evidence that any other specific features of the testing application may influence the strength of the test-criterion relationship, the correspondence between those features in the local situation and in the metaanalysis should be reported. Any significant disparities that might limit the applicability of the metaanalysis findings to the local situation should be noted explicitly.
1.23	p. 30	Any metaanalytic evidence used to support an intended test score interpretation for a given use should be clearly described, including methodological choices in identifying and coding studies, correcting for artifacts, and examining potential moderator variables. Assumptions made in correcting artifacts such as criterion unreliability and range restriction should be presented, and the consequences of these assumptions made clear.
1.24	p. 31	If a test is recommended for use in assigning persons to alternative treatments, and if outcomes from those treatments can reasonably be compared on a common criterion, then, whenever feasible, supporting evidence of differential outcomes should be provided.

Note. [1]In American Educational Research Association et al. (2014).

Table 70.
Standards for Cluster 3(f) Consequences of Testing.

Standard	Ref.[1]	Standard Text
1.25	p. 30-31	When unintended consequences result from test use, an attempt should be made to investigate whether such consequences arise from the test's sensitivity to characteristics other than those it is intended to assess or from the test's failure to fully represent the intended construct.

Note. [1]In American Educational Research Association et al. (2014).