OPTIMIZING THE EFFICIENCY AND PRECISION OF THE DELAWARE SCHOOL CLIMATE SCALE: AN APPLICATION OF MULTIVARIATE GENERALIZABILITY THEORY

by

Dandan Chen

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Education

Spring 2019

OPTIMIZING THE EFFICIENCY AND PRECISION OF THE

DELAWARE SCHOOL CLIMATE SCALE:

AN APPLICATION OF MULTIVARIATE GENERALIZABILITY THEORY

by

Dandan Chen

Approved: ____

Chrystalla Mouza, Ed.D.
Director of the School of Education

Approved: ____

Carol Vukelich, Ph.D.
Dean of the College of Education and Human Development

Approved: ____

Douglas Doren, Ph.D.
Interim Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Henry May, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

George Bear, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Joseph Glutting, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Jonathan Rubright, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I would like to gratefully acknowledge the guidance, support and encouragement of my doctoral advisor, Dr. May, and my secondary advisor, Dr. Bear, as well as my committee members, Dr. Glutting and Dr. Rubright.

I would like to extend my sincere thanks to the faculties and staff in the School of Education at the University of Delaware for guiding me through the journey of my graduate program. My gratitude extends to the faculties and staff at the Department of Psychology at South China Normal University for leading me to the theories and research in the fields of psychology and education.

Finally, I would like to express my deepest appreciation to my grandma (Laolao), my parents, my husband, the rest of my family, and friends who believed I could become Dr. Chen long before I did.

TABLE OF CONTENTS

Appendix

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This dissertation examined issues about the efficiency and precision of Delaware School Climate Scale (DSCS) scores in elementary, middle, and high schools, and among samples of students, teachers, and parents across three years. It also contributed to the nascent research on measuring school climate as a multilevel, multi-informant (i.e., student, teacher, parent), and multi-dimensional construct using survey questionnaires. To illustrate how multivariate generalizability theory can address issues in measuring school climate, I rely on data from the DSSC to demonstrate the application of the proposed approach. In total, students in elementary (n = 41,754), middle (n = 32,919), and high schools (n = 19,374) responded to the DSCS-Student survey composed of 30 items in seven subscales. Teachers in elementary (n = 5,542), middle (n = 2,885), and high schools (n = 2,577) responded to the DSCS-Teacher/Staff survey composed of 38 items in nine subscales. Parents in elementary (n = 27,310), middle (n = 8,641), and high schools (n = 2768) responded to the DSCS-Home survey composed of 25 items in six subscales. A multivariate generalizability theory model with the class-means design was used to investigate the sources of variation in the DSCS scale and subscale scores: person within school, items, and interactions among those facets and the object of measurement, schools. Separate G and D studies were conducted, and school profile scores were produced for student, teacher, and parent groups by grade levels and years. Then, a second-order model was used to examine the sources of variation when measuring school climate longitudinally as a multi-informant index: occasions and interaction between

occasions and the object of measurement, schools. The results of these two models indicated that the DSCS can provide reliable and precise estimates of school-level school climate in most groups, except for the parent respondent group for elementary schools across years. The results of D studies also showed that fewer respondents per school are needed to reach the recommended threshold of .80 of G coefficients in most groups.

Specific recommendations for future DSCS measurement procedures and implications for the use of the DSCS scale and subscale scores are provided. Implications for measuring school climate and sampling respondents are also discussed. If the future measurement procedure focuses on students, fewer students (40% of the original number of students per school) and one less item per subscale is necessary to reach the recommended threshold .80 across grade levels. For the teachers' sample, fewer teachers per school (60% of the original number of teachers per school) are needed to achieve the threshold of reliability in middle and high schools. For the parents' sample, fewer parents per school (80% of the original number of parents per school) are required to achieve the recommended threshold in high schools, but more parents per school are required in elementary and middle school. In addition, the results suggest that the use of multiple survey forms, with matrix sampling of items, can simultaneously increase reliability while also reducing response burden. If resources are limited, a survey of students or teachers can be sufficient for evaluating school-level school climate.

Chapter 1

BACKGROUND

1.1 Introduction

Research regarding the role of school climate in promoting safe and caring learning environments has a long and rich history. This interest has been fueled in recent decades by findings from prospective studies that link the contextual effects of school climate to school outcomes, such as school-wide bullying, academic achievement, and teacher burnout (Towles-Reeves, et al., 2012). Furthermore, education ministries around the world have supported researchers and educators in developing evidence-based approaches to promote positive school climate (Thapa, Cohen, Guffey, & Higgins-D'Alessandro, 2013). Perhaps more than ever before, the climate and safety of our schools are being questioned and evaluated. Although there has recently been an increased focus on school climate in education practices and research, there are still concerns about how school climate can be evaluated reliably and comprehensively (Huang, F. L. & Cornell, 2015; Zullig et al., 2015) particularly regarding multi-level structures in school climate measurements.

This dissertation is built on both conceptual and empirical arguments about how best to measure and scale indicators of school climate. To support these arguments, this dissertation includes a review of prior research on school climate and a comparison of the scaling methods currently used for the Delaware School Climate Scale (DSSC) with methods based on a generalizability theory (G theory) approach in relation to each of the following issues: (1) Nesting of individual respondents within organizations (i.e., with organizations being the focal unit of analysis); (2) Handling data from multiple respondent groups; (3) Handling longitudinal data at item and respondent levels. Consequently, the primary goal of my dissertation is to demonstrate increased precision of DSSC based upon G theory as well as offer guidance for increasing efficiency in future sampling designs.

1.2 Organization of This Dissertation

To address these issues, the background of school climate research and its measurement are discussed in the rest of this chapter. In Chapter 2, a brief review of theory and research on school climate is presented, and the concepts and methods of generalizability theory are reviewed. In Chapter 3, primary research questions are presented, and the methodology used in this dissertation are described in detail, including research design, data sources, and analytic approaches. In Chapters 4 and 5, results from multivariate generalizability theory analyses are presented and discussed.

1.3 Why Measure School Climate?

The concept of school climate, "the quality and character of school life," originated from organizational climate and school effectiveness research (Cohen, McCabe, Michelli, & Pickeral, 2009). With the movement of policy and reform in education, school climate has become a measurable index of school psychological quality perceived by teachers, students, and parents. What are the rationales behind the prevalent use of school climate measures? What are the goals for measuring school climate? How should researchers, policymakers, and practitioners use school climate data? All these questions are related to the purposes of measuring school climate.

The purposes of measuring school climate include providing a form of feedback to teachers, students, and parents as well as informing policy makers about the quality of the school (Huang, F. L. & Cornell, 2015; Schweig, 2014). Teachers and principals can use school climate data to evaluate and assess the collective perceptions of school structure and support in a school. As a result, data from school climate surveys can be used as a self-evaluation tool, as well as provide parents with comprehensive and quantified feedback. In addition, the topic of school climate has received significant attention from the education policy community. Almost half of the states in the U.S. have included school climate as a key element of educational policies, and some states have even integrated school climate into their school improvement standards (Piscatelli & Lee, 2011). School climate measures are also widely used in education research studies because the term itself implies the quality of school environments. However, several empirical studies have argued that school climate should be considered not just as a school-level phenomenon, but within a multilevel framework in which individual respondents are nested within schools and multilevel modeling analysis is used to explore climate indicators at the school level as well as within schools (Huang, F. L. & Cornell, 2015; Konold & Cornell, 2015).

Another purpose of measuring school climate is to provide an outcome measure for evaluating the impact of school interventions (e.g. Positive Behavioral Support programs) and predictors for policy relevant evaluations (Schweig, 2014). Research on school reform and school discipline has identified positive school climate as a potential target of interventions because school climate is directly associated with bullying, student engagement, positive youth development, and academic achievement (Cohen et al., 2009; Gage, Prykanowski, & Larson, 2014). Also, the predictive function of school climate has been documented in several evaluation studies. For example, an evaluation report from the Bill and Melinda Gates Foundation found that student success is influenced by a supportive and caring environment, in conjunction with rigorous learning experiences (Shear et al., 2008). Similarly, a recent evaluation study involving a randomized control trial found a positive impact of a leadership program promoting positive school climate and improving student achievement (Jacob, Goddard, Kim, Miller, & Goddard, 2015). These findings emphasize the usefulness and meaningfulness of school climate as an indicator of the success of interventions in education settings.

Beyond the practical purposes of measuring school climate, there is also a need to build theories and develop measures that concentrate on the relationship of school climate with aggression and bullying (Espelage, Low, & Jimerson, 2014). Through best practices grounded in sound theories, researchers in the field have noted the importance of multilevel structures of school climate and considered multilevel statistical methods to understand the relationships among school climate, bullying, and discipline (Huang, F. L. & Cornell, 2015; Konold et al., 2014). However, only a few studies have extended school climate research from cross-sectional studies and individual-level instruments to longitudinal studies and multilevel instruments. For example, two longitudinal studies reported how school climate corresponded with changes in bullying perpetration, aggression, and school safety (Gage et al., 2014; Turner, Reynolds, Lee, Subasic, & Bromhead, 2014). Likewise, Huang and Cornell (2015) pointed out that few instruments account for the multilevel nature of school climate data. They demonstrated how their instrument supports a conceptual framework of school climate using multilevel factor analysis. These studies confirm the need to develop multilevel instruments in order to conduct rigorous research considering the theoretical structure and influence of school climate.

1.4 The Problem of Measuring School Climate

The concept of school climate is generally considered as “patterns of people’s experiences of school life that reflects norms, goals, values, interpersonal relationships, teaching and learning practices, and organizational structures” (Cohen, et al, 2009). This definition also suggests four dimensions of school climate: safety, teaching and learning, relationships, and organizational structures. With each domain, multiple constructs and indicators are possible. The safety domain emphasizes the degree to which individuals feel physically and social-emotionally safe in school. The teaching and learning domain represents the quality of instruction (e.g. high expectations for academic achievement), social, emotional and ethical learning, on-going professional development, and effective leadership in school. The relationship domain is composed of respect for diversity, school community, and school connectedness. The organizational or environmental-structural domain refers to the physical environment, school size, adequate recourses, and availability of supplies.

Due to the broad conceptual models of measuring school climate, there is little consensus on the construction of school climate instruments. Table 1 lists a series of psychometrically sound school climate measures that have been published in peer-reviewed journals. These measures comprise constructs from multiple domains of school climate, including teacher-student relations, student-student relations, respect

Table 1.1: Summary table of school slimate instruments

Measure and authors	Theory and model	Respondent	Psychometric analysis	Local program
Delaware School Climate Survey(Bear, Yang, & Pasipanodya, 2014; Bear, Gaskins, Blank, & Chen, 2011; Bear, Yang, Pell, & Gaskins, 2014)	(a) Authoritative discipline theory (Baumrind, 1971, 1996; Bear, 2005) and (b) Stockard and Mayberry's (1992) theoretical framework of school climate	Student, Grade 3-12 Teacher, Grade 3-12 Parent, Grade 3-12	EFA & CFA, measurement invariance testing among race, gender, and grade levels Multilevel analysis, group-mean centering	A combination of SWPBS and SEL
Authoritative School Climate Survey (Huang et al., 2015; Konold et al., 2014)	Authoritative discipline theory	Student, Grade 7 Teacher, Grade 7-8	EFA Multi-level CFA Multi-informant of gender	The state's annual School Safety Audit program
Student and parent questionnaire (Griffith, 1995; Griffith, 2000)	Stockard and Mayberry (1992) Moos's (1979) relationship-growth aspects of school social climate	Student, Elementary schools Parent, Elementary schools	EFA	
Maryland Safe and Supportive Schools Climate Survey (Bradshaw, Waasdorp, Debnam, & Johnson, 2014)	The US Department of Education (USDOE) developed a 3-factor model of school climate comprised of safety, engagement, and environment.	Student, Grade 9-12	EFA & CFA, measurement invariance about student sex, grade level, and ethnicity.	SWPBS
School Climate Survey (Zullig et al., 2014; Zullig, Koopman, Patton, & Ubbes, 2010; Zullig et al., 2015)	Not mentioned, but explained five theorized domains/ Organizational Change Theory (Kaluzny & Hernandez, 1988)	Student, Grade 6-12/Grade 9-12	EFA & CFA	
Inventory of School Climate (Brand, Felner, Shim, Seitsinger, & Dumas, 2003; Brand, Felner, Seitsinger, Burns, & Bolton, 2008)	Moos (1979) and his colleagues proposed three overarching dimensions of social climate	Student, Grade 6-8 Teacher, Grade 6-8	EFA & CFA	
Developmental Study Center's School Climate Survey (Ding, Liu, & Berkowitz, 2011; Liu, Ding, Berkowitz, & Bier, 2014)	Social-cognitive theory (Bandura, 2001)	Student, Grade 6-10 Teacher, Elementary school	EFA & CFA Longitudinal analysis	

Table 1 continued

The California School Climate and Safety Survey& Brief version(Furlong et al., 2005; You, O'Malley, & Furlong, 2014)	These most widely used school safety instruments were originally developed to assist in the creation and monitoring of national trends of school violence using a public health model.	Student, Grade 6-12 Teacher, Grade k-12	EFA & CFA, measurement invariance across grade levels	the CSCSS was designed to correspond with the school safety-planning model in California.
The Georgia Brief School Climate Inventory (GaBSCI) (White, La Salle, Ashby, & Meyers, 2014)	Not mentioned.	Student, Grade 6-8	EFA & CFA	

for diversity, school safety, clarity of expectations, fairness of school rules (or discipline), bullying in schools, and physical environment. Because school climate is a broad concept with multiple domains and sub-domains, each of these instruments intends to assess most of the school climate domains, especially relationships and safety in schools. Although a number of psychometric validations have been done, the ways of measuring school climate are not without limitations.

The most critical limitations are the conception of and methods for measuring school climate in a multilevel framework. Conceptually, some constructs of school climate might be more appropriately interpreted at the organizational level and others might be considered as an individual level phenomenon (Anderson, 1982). For example, Konold and his colleagues (2014) first attempted to measure the structure of school climate at both school and student levels and found that the application of a school climate score could be either student or school level (i.e., except for the Cognitive Engagement Scale, which was primarily a school level construct). However, in their study, items measured individuals' perceptions rather than focusing on overall school conditions.

Many studies have raised concerns regarding the multilevel nature of data structures in measuring and analyzing school climate (Bear, Gaskins, Blank, & Chen, 2011; Huang, F. L., 2014). In most situations, the factor structures of school climate instruments were examined at the individual level while the analysis of the relationship between school climate and student outcomes were conducted at both the student and school levels. Inconsistency in the focal unit of the analysis might cause problematic interpretations of the results and their implications. In industrial-organizational psychology, a number of studies have employed G theory to analyze the relationships between variables at both the individual and organizational levels (Epitropaki & Martin, 2004; Hoffman, Olson, & Haase, 2001). In this dissertation, G theory is used to handle the multilevel structure of school climate.

In addition, greater attention has been paid to the 360-degree multiple rater approach of measuring school and student outcomes, including the topic of school climate. Different raters' perceptions cannot be separated from the social environment because of the dynamic interactions between persons. A growing number of studies have separately examined the factor structure of students, teachers, and parents' reports on school climate instruments and attempted to use a multi-informant instrument as a comprehensive indicator of school climate (Bear, Yang, Pell, & Gaskins, 2014; Konold & Cornell, 2015). However, these studies have only tested measurement invariance across rater groups rather than the variance of school climate structures across different rater groups.

Although these studies have not provided a practical analytic approach of evaluating different raters' perceptions of school climate, several studies focusing on behavioral and academic assessments have shed light on the procedures for analyzing data from 360 degree, multiple-rater assessments within a G theory framework, which considers rater variance accounting for the variance of a multiple-facet construct

(Briesch, Chafouleas, & Riley-Tillman, 2010; Hoffman, Olson, & Haase, 2001). In addition, G theory can also handle the interactions between different measurement factors, such as rater groups and grade levels. Although a number of studies have shown that students in different grade levels perceive school climate differently due to their psychological development (Wang & Dishion, 2012; Way, Reddy, & Rhodes, 2007), the cross-rater group relationships between students, teachers, and parents' perceptions of school climate across grade levels are still unknown. If researchers are to develop better measures of school climate, it will be necessary to account for and understand all major sources of variance.

Finally, the solutions to temporal issues of measuring school are unclear: whether perceptions of school climate change as students grow older, whether school climate as an organizational construct changes over time, and how should we measure these changes. On one hand, several longitudinal studies found declines in early adolescents' perceived school climate (Roeser, Eccles, & Sameroff, 2000; Way, Reddy, & Rhodes, 2007). In addition, a number of cross-sectional studies confirmed that middle school students reported the less favorable perceptions of school climate than elementary and high school students (e.g., Yang, et al., 2013). On the other hand, without interventions or school reforms, students' and teachers' perceptions of school climate are quite stable (Ding & Hall, 2007; Liu, Ding, Berkowitz, & Bier, 2014). If we are not clear about the trajectory of perceived school climate along the stages of development, temporal issues become significant obstacles to research design in experiments and intervention studies intended to capture organizational-level changes in climate.

1.5 Need for Generalizability Theory

The process of measuring school climate can be summarized as a multi-dimensional, multi-level, multi-informant, and longitudinal measurement procedure. These aspects of measuring school climate can be treated as different sources of variation in a generalizability theory framework. Generalizability theory extends Classical Testing Theory (CTT) by using ANOVA to disentangle the undifferentiated error term in CTT to multiple sources of error. In univariate G theory, a single universe score is defined in the universe of admissible observations. In multivariate G theory, multiple universe scores and their composite exist in the parallels of universes (Brennan, 2001). As such, multivariate G theory provides flexible and robust tools for analyzing complicated measurement procedures, such as school climate assessment.

The issues of measuring school climate can be treated as different sources of error contributing to the scale and subscale scores of school climate measures. The multi-dimensional aspect of measuring school climate can be considered as the linked facet in multivariate G theory framework that connects items within each subscale to the composite score of the total scale. The multi-level aspect can be treated as the school facet, the object of measurement, and the person within school facet. In other words, it is equivalent to the between and within school variations in the scale and subscale scores of school climate measures. The multi-informant aspect can be considered as a rater-group (i.e., students, teachers, and parents) facet attributable to the homogeneous perceptions of the climate in a school. The longitudinal aspect can be treated as an occasion facet that contributes variations to the stability of school climate scores. To sum up, multivariate generalizability theory provides an appropriate analytic model for disentangling variation in total scores into components associated with school, person within school, item, subscale, rater, and occasions in the process of measuring school climate. By determining which component contributes most to the imprecision of measurement, subsequent procedures can be designed to reduce those measurement errors and optimize the generalizability of school climate scores.

To illustrate how multivariate G theory can address issues in measuring school climate, I rely on data from the Delaware School Climate Scale (DSCS) to demonstrate the application of the proposed approach. The rationales are (a) the DSCS is one of the few psychometrically sound measures of school climate that have been published in peer-reviewed journals and the datasets of the DSCS are available for secondary analysis; (b) the DSCS provides a global view of three different groups of raters across all grade levels; and (b) the DSCS has been administered in Delaware public schools for seven years (beginning in 2010). In other words, the measurement design and evaluation of the DSCS allows G theory to explore the relative variability across individuals, rater groups, grade levels, and over time. The next chapter reviews the related theories and measures of school climate, explain the scaling methods and psychometric issues of the measures, and multivariate G theory models in estimating variance components of a measure.

Chapter 2

LITERATURE REVIEW

This chapter describes literature relevant to school climate as the focus of this dissertation. It is organized into five sections: (1) a brief review of the theoretical foundation of school climate, (2) an illustration of two conceptual frameworks of school climate, (3) a presentation of the DSCS measurement model, (4) a discussion of methodological issues in measuring school climate, and (5) a brief introduction of G theory.

2.1 Theoretical Foundations of School Climate: A Brief Review

As early as 1908, Perry (1999) addressed the concept of school climate in his book on school management. He noted the important claim regarding the school as an organization that benefits students' moral development through school spirit. Later, with the movement of field theory in social psychology (Lewin, 1939) scholars focused with increased attention on the impact of human environments on individual behavior and with less interest on how personal traits influence individual behavior (Bronfenbrenner & Morris, 2006).

Since then, there have been a number of empirical and theoretical studies assessing how child outcomes are affected by environmental variables such as school leadership atmosphere, classroom interaction, neighbor, and families. These studies, though very different, contribute to the modern theoretical framework of school climate by emphasizing the effect of the environment. For instance, the pioneering work of the Organizational Climate Description Questionnaire (OCDQ) examined the interactions between teacher and school administrators in an elementary school (Halpin & Croft, 1963). A measurement model of the Classroom Environment Scale delineated nine types of classroom climates from students' perspectives (Moos, 1973). Furthermore, the ecological model of human development and its most recent iteration provide a theoretically sound framework in developmental psychology and education policy research (Bronfenbrenner, 2010), while the authoritative discipline theory reveals critical insights into parents and educators' discipline practices (Baumrind, 1966; Baumrind, 1996). Even though the authoritative discipline theory model focuses on parenting styles and child development, the theory has also been applied to school discipline and school climate (Bear, 2010). These models and research serve the foundations of contemporary school climate research. The following will review the aforementioned frameworks and discuss the connections with contemporary theories and instruments.

2.1.1 School Climate as an Organizational Phenomenon

The Organizational Climate Description Questionnaire constructed by Halpin and Croft (1963) is generally recognized as the first scientific attempt to characterize the organizational climates of a school (Stockard & Mayberry, 1992). The measurement construction of the OCDQ is similar to personality assessment because it views the climate within an elementary school as the "personality" within an individual (Anderson, 1982; Halpin & Croft, 1963). The OCDQ was composed of 64 Likert-scale items. The results of exploratory factor analysis classified four characteristics of teachers' behaviors – "disengagement, hindrance, and esprit, intimacy" – and four types of principals' behaviors – "aloofness, production emphasis, thrust, consideration." The questionnaire also used profile analysis to categorize the six types of school climates: the open climate, the autonomous climate, the controlled climate, the familiar climate, the paternal climate, and the closed climate. The profile scores of the six types of school climates characterized three patterns of social interactions in schools, including authenticity, satisfaction, and leadership initiation (Halpin & Croft, 1963).

The interaction between teachers and principals is the key assumption of the OCDQ. It determines the variabilities of school climate and establishes the foundation for assessing it through teachers' and principals' perspectives in school effectiveness and administration research (Hoy & Hannum, 1997). This assumption and the measurement model of the OCDQ provide practical utility and theoretical justifications for categorizing organizational climates. As a result, the questionnaire plays an important role in promoting studies on organizational climates in schools or other types of organizations, such as business organizations and college working climates (Anderson, 1982).

Subsequent research has continued to revise the questionnaire, keeping it up to date. The OCDQ has been revised and adjusted several times. The first revision of the OCDQ was designed to define the climate of an elementary school as a healthy working environment for teachers and administrators (Hoy & Clover, 1986). Later, the survey was refined for assessing middle school climates focusing on organizational health and student achievement (Hoy & Hannum, 1997; Kottkamp, Mulhern, & Hoy, 1987). The original questionnaire and revision of the OCDQ still have a profound impact on contemporary school climate measurement in evaluating teachers' and administrators' perceptions of organizational climate (You, O'Malley, & Furlong,

2014) as well as students' perceptions of a positive environment in school (Liu et al., 2014; Zullig, Koopman, Patton, & Ubbes, 2010).

2.1.2 Classroom Environment and School Climate

The Classroom Environment Scale is another early scientific assessment of school climate that measures middle and high school teachers' and student's perceptions of classroom climate. In his review of assessing and characterizing human environments, Moos (1973) noted that "substantial proportions of the variance in behavior are accounted for by situational and environmental variables". The measurement model of the scale offers a theoretically sound basis for the development of school climate understandings. The model recognizes relationships, personal growth, and system maintenance as the three domains of a healthy and positive classroom climate. To date, a number of school climate instruments have employed personal relationships and system maintenance as the main domains of their measurement models, such as the Inventory of School Climate (Brand, Felner, Shim, Seitsinger, & Dumas, 2003), the Maryland Safe and Supportive Schools Climate Survey (Bradshaw, Waasdorp, Debnam, & Johnson, 2014), the Delaware Surveys of School Climate (Bear et al., 2011), and the School Climate Questionnaire (Griffith, J., 1995; Griffith, James, 2000).

2.1.3 School Climate from an Ecological Perspective

The ecological theory of human development emphasizes the values of "the context of child rearing," including the interactions among child, family, school, neighbor, and other social systems on the well-beings of children and societies (Bronfenbrenner, 1975). This theory "focuses attention on development as a function of interaction between the developing organism and the enduring environments or contexts in which it lives out its life" (p429, Bronfenbrenner, 1975). A child's life is not isolated from the aforementioned social systems. The interactions of these social systems are likely to determine children's academic achievement, social-emotional learning, and engagement in schools. Bronfenbrenner's theory has greatly influenced contemporary studies on school effectiveness and school climate research. For example, several studies found that high-quality relationships between students, parents and teachers were associated with students' improved achievement (Bryk, Sebring, Allensworth, Easton, & Luppescu, 2010).

Bronfenbrenner kept reassessing and revising the ecological model up until the 2000s. He refined it as a bioecological model emphasizing four primary factors: process, person, context, and time (PPCT), as well as the interactions between these factors (Bronfenbrenner & Morris, 2006). On one hand, this model employs a comprehensive system to explain how children and environments change overtime. It also provides general theoretical supports for using multi-informant and multi-level assessment (Huang, F. L. et al., 2015; Koth, Bradshaw, & Leaf, 2008) to longitudinally evaluate the effects of school climate (Thapa et al., 2013). On the other hand, it seems difficult for contemporary school climate measurements to capture all the primary factors of the PPCT model due to its complexity.

2.1.4 Discipline and School Climate

The authoritative model of parenting (Baumrind, 1966; Baumrind, 1996; Baumrind, 2013) influences the development of contemporary school discipline and school climate research. Although the core of the model is the typology of parenting, the ultimate objectives of parental discipline and school discipline are similar and aim to achieve both the short-term goals of obedience and compliance as well as the long term goals of self-discipline and autonomy (Bear, 2010a). The authoritative model proposed by Baumrind in the 1960s identifies *authoritative parenting* as the midpoint between authoritarian (conservative) and permissive (liberal) parenting. These three parenting styles were classified by two primary elements of childrearing: responsiveness and demandingness. Responsiveness refered as “the extent to which parents intentionally foster individuality and self-assertion by being attuned, supportive, and acquiescent to children’s needs and demands” (Baumrind, 1996). Demandingness refers to “the claims that parents make on children to become integrated into family and community by their maturity expectations, supervisions, disciplinary efforts, and willingness to confront a disputative child” (Baumrind, 1996). The authoritative parenting is a combination of high responsiveness and demandingness. The authoritarian parenting style is extremely demanding but not responsive, while a permissive parenting style is extremely responsive but not demanding (Baumrind, 1996). Maccoby and Martin (Maccoby & Martin, 1983, cited in (Pellerin, 2005) expanded Baumrind’s typology of parenting by placing responsiveness and demandingness into a two by two table and adding neglecting parenting which is neither demanding nor responsive.

Ample research shows that the authoritative parenting style, with a balanced integration of responsiveness and demandingness, tends to predict high academic achievement, high social competence, and few behavior problems (Gregory, Anne & Cornell, 2009; Lamborn, Mounts, Steinberg, & Dornbusch, 1991). This parenting style is beneficial to the achievement of children from different racial and ethnic groups. Baumrind’s early work found that authoritative parenting has a positive impact on

Caucasian children’s achievement(Baumrind, 1996). Later on, several studies have also found the positive impact of authoritative parenting on the achievement of African American children (Darensbourg & Blake, 2014; Taylor, Hinton, & Wilson, 1995). A landmark study by Lamborn and his colleagues (1991) demonstrated that adolescents with authoritative parents reported higher scores of psychological competence and lower scores of psychological and behavioral dysfunctions than those students with authoritarian, permissive, or neglecting parents. Likewise, Steinberg and his colleagues found that adolescents with authoritative parents were more engaged in school activities (Steinberg, Lamborn, Dornbusch, & Darling, 1992). Collectively, these studies highlight the effectiveness of authoritative parenting on the children’s developmental outcomes.

In school climate research, responsiveness is often referred to as *social support* and demandingness as *structure* (Bear, 2010a; Gregory, Anne & Cornell, 2009). Support is the extent to which teachers and school administrators meet students’ social and developmental needs. Structure is defined as the extent to which teachers articulate high behavioral expectations and fair rules, implement rules consistently, and manage students’ behavior with support (Bear et al., 2011). A number of studies have indicated that authoritative school climates with positive support and structure produce the best results on student engagement and behavioral outcomes (Gill, Ashton, & Algina, 2004; Gregory, A. et al., 2010; Lee, Olszewski-Kubilius, & Thomson, 2012; Pellerin, 2005). For instance, Grill et al., (2004) found that students’ perceptions of support predicted their engagement and internal control. Pellerin (2005) demonstrated that authoritative schools had the most influential effect on student disengagement (i.e. students’ self-reports of lateness and class-cutting) than authoritarian, permissive, and neglecting schools. Gregory et al (2010) found that an authoritative school climate related to fewer cases of bullying.

Taken together, the principles of the authoritative parenting discipline theory also apply to school climate research. Schools with high social support and structure tend to have more favorable perceptions of school climate from all members of the school community. By contrast, an unbalanced combination of social supports and structure is less likely to provide a nourishing and warm school climate.

2.2 Two Modern Theoretical Frameworks of School Climate

Over the last two decades, researchers have constructed and developed school climate instruments based on the aforementioned theoretical frameworks and the needs of local education systems. As summarized in Table 1, almost all of the school climate surveys were supported by state or federal school improvement plans or climate related programs. In other words, in order to meet the needs of local agencies, researchers adjusted the measurement model of the instruments to the goals of the prevention programs or the school improvement plans. As a result, the US Department of Education Safe and Supportive Schools (USDOE S3) model was widely recognized as providing the guiding principles and framework for constructing school climate instruments. Another widely recognized measurement model of school climate instruments is the authoritative discipline model because of its effectiveness in promoting warm and safe school climate. Thus, the following sections will focus on the USDOE S3 model and the authoritative discipline model.

2.2.1 The US Department of Education (USDOE) Model of School Climate

Recognizing the new appreciation of school climate research, the U.S. Department of Education's Office of Safe and Drug Free Schools (2009) designed the Safe and Supportive Schools (S3) programs and awarded 11 states with grants to conduct statewide school climate assessments and evaluation systems. The S3 grant programs stimulated substantial interest in measuring school climate and implementing prevention programs that aim to improve it (Konold et al., 2014). Among the grantees of the S3 initiatives, the California School Climate, Health, and Learning Survey (Cal-SCHLS), the School Climate Measure (Zullig et al., 2015), and the Maryland Safe and Supportive Schools Climate Survey (MDS3) (Bradshaw et al., 2014) provided appropriate evidence of high psychometric qualities, and the results of the latter two surveys were published in peer-reviewed journals. In addition, other surveys that had undergone the peer-reviewed process aligned the overarching structure of the surveys with the USDOE S3 model of school climate, such as the Brief-California School Climate Survey (You et al., 2014) and the Authoritative School Climate Survey (Konold et al., 2014).

As discussed, there is no universal definition of school climate; the USDOE S3 model depicts a big umbrella of all the factors related to the subject. The model explains that a positive school climate is associated with engagement (e.g. strong relationships, respect for diversity in schools, and school connections with communities), safety (e.g. emotional and physical safety and substance use) and environment (e.g. healthy physical, academic, and disciplinary environment and wellness) (U.S. Department of Education, 2013, cited in Bradshaw et.al., 2014). Among the aforementioned surveys, the MDS3 high school student survey closely examined the 3-factor USDOE S3 model, including 13 subscales: engagement (connection to teachers, student connectedness, academic engagement, school connectedness, equity, and parent engagement), safety (perceived safety, bullying and aggression, and drug use) and environment (rules and consequences, physical comfort and support, disorder). Bradshaw and her colleagues also conducted several studies that supported the structure of the USDOE S3 model, showing that the engagement domain, especially teacher and student connectedness, was associated with adolescents' adjustment problems across gender and grade level (Hurd, Hussain, & Bradshaw, 2015; Morin, Bradshaw, & Berg, 2015).

2.2.2 Authoritative Discipline Model of School Climate

Authoritative discipline theory considers the authoritative discipline method to be the most effective discipline style, which is composed of a healthy balance of support and structure. This assumption is supported by research on childrearing (Baumrind, 1996) and on school discipline and school climate (Bear, 2010a; Gill et al., 2004; Gregory et al., 2010; Pellerin, 2005). A balanced combination of support and structure fosters self-discipline in students and promotes a healthy school climate demonstrated by positive interpersonal relations, safety, prosocial behavior and limited misbehavior (Gregory et al., 2010). There are several explanations for why the authoritative discipline model of school climate works. First, psychological autonomy and behavioral compliance are considered two interdependent concepts in the model, which eliminates teachers' misunderstanding of using discipline to manage student misbehavior. Second, the model distinguishes the differences between authoritative and authoritarian teachers. An authoritative teacher uses discipline in an informative style, while an authoritarian teacher employs it in an assertive way. Third, positive support and organized structure satisfy teachers' needs of effective classroom management and students' needs of being cared for and loved. Thus, a positive and healthy school climate traces back to authoritative discipline styles.

The authoritative discipline model offers a theoretically sound basis for school climate measurement. For example, the main theoretical roots of the Delaware School Climate Scale (DSCS; Bear, et al., 2013) and the Authoritative School Climate Survey (ASCS; Huang, et al., 2015) are the authoritative discipline model of school climate. These surveys provide some of the most notable school climate measurements with high reliability and validity. Each of these surveys facilitates the implementation of prevention programs in local schools and provides an evaluation outcome examining the effectiveness of these programs. For instance, the DSCS is used to investigate the implementation and effectiveness of Delaware's initiative to improve its school climate, which integrates the School-Wide Positive Behavior Supports and Social and Emotional Learning approaches (Bear et al., 2011). The measurement model of ASCS was employed by the professional development program, My Teaching Partner-Secondary (MTP-S), which provides coaching assistance for teachers to engage in positive interactions with students (Gregory, Anne, Allen, Mikami, Hafen, & Pianta, 2014).

Another contemporary theoretical framework of school climate proposed by Stockard and Mayberry (1992) emphasizes similar conceptualization as the authoritative discipline theory. They conducted a comprehensive review of psychological, sociological, economic theories and empirical studies on school effectiveness and school climate. They created two broad terms, social action and social order, which are similar to responsiveness and demandingness in Baumrind's theory. Social action represents daily social interactions among students, teachers and staff in a caring and supportive manner. Social order emphasizes the norms, values and organizational structures to maintain a safe environment. The findings of several studies supported Stockard and Mayberry's framework of school climate, which indicates positive associations between school climate and academic performance, as well as satisfaction at the individual level (Griffith, 1995; Griffith, 1997) and achievement and engagement at the school level (Bear et al., 2011).

2.3 The Measurement Model of the DSCS

The theoretical roots of the DSCS are the aforementioned authoritative discipline theory (Baumrind, 1996; Gregory & Cornell, 2009) and Stockyard and Mayberry's (1992) framework on school climate. The measurement model of the DSCS is also guided by the social-ecological perspective that individuals' perceptions of social systems play a critical role in understanding human behavior rather than objective reality (Bronfenbrenner, 1979). The variation of individuals' perceptions occurs even in the same social system. For example, in a school with an authoritarian discipline style, teachers may perceive a strong and organized school climate, while students may be unsatisfied due to a perceived lack of empathy, believing the teacher to be too demanding or overbearing. Authoritarian teachers employ pervasive discipline to manage students, but provide few social supports for them (such as ignoring students' questions when they have difficulties), perhaps leading students to perceive school climate as a cold and uncaring atmosphere. In sum, individuals' perceptions are generally due to their own experiences rather than objective observation. The social-ecological perspective highlights the need for measuring the perceptions of all members of a school community.

Following to its theoretical roots, the DSCS evaluates the two dimensions of school climate, *social support and structure*, which align with the aims of the local prevention programs in each state. Under the support dimension, there are three subscales including teacher-student relations and student-student relations. Under the structure dimension, there are three subscales including clarity of behavioral expectations, school safety, and fairness of school rules. The DSCS also included Student Engagement School-Wide and Bullying School-Wide subscales on student and teacher versions due to the emphasis that recent research has placed on them as contributing factors to school climate (Cohen et al., 2009; Thapa et al., 2013). Students, teachers and parents report the degree to which they agree to an item in every subscale. In order to allow comparisons between different grade levels, the survey was designed to measure respondents from elementary (from third grade), middle, and high schools (Bear et al., 2016).

These subscales have also been theorized and are found in other psychometrically sound school climate instruments. For example, Teacher-Student Relations, Student-Student Relations (called Peer Interactions), Clarity of Behavioral Expectations, School Safety and Respect for Diversity (called Support for Cultural Pluralism) are found in the Inventory of School Climate (Brand et al., 2003), School Climate Survey (Zullig et al., 2010; Zullig et al., 2015), Developmental Study Center's School Climate Survey (Liu et al., 2014) and other school climate instruments but with different subscale names. However, these subscales are commonly found in students' surveys rather than teachers' or parents' surveys, which hinders comparisons between different respondent groups. Students, parents and teachers are important members of a school community; different informants provide distinct and unique perceptions of school climate. As a result, multi-informant school climate instruments (i.e. measuring the same constructs from different rater groups), such as the DSCS, may provide a more comprehensive and diverse view of the extent to which school members perceive school climate (Bear et al., 2014).

2.3.1 Research Supporting the Dimensions of the DSCS

A variety of research findings have indicated the importance of two dimensions of school climate, *social support and structur*e, in measuring and evaluating school climate (Bear et al., 2011; Gill et al., 2004). The early work of authoritative school climate research applied Baumrind's authoritative discipline theory to school settings and categorized four types of school climate as parenting styles in Baumrind's theory. For instance, Gill and his colleagues (2004) used a seven-item scale of students' perceptions of school responsiveness (social support) to examine the effects of authoritative school climate on student engagement. Students who held more favorable perceptions of social supports reported higher scores of engagement and internal control. The schools with higher social support tended to have a smaller achievement gap in mathematics. Using the High School Effectiveness Study dataset, Pellerin (2005) examined which types of schools predict student disengagement and found that schools with high social support and structure had a positive impact on student disengagement. In these studies, the school climate instruments were a part of national surveys that consisted of only several items, which may not provide sufficient and reliable information. More recently, researchers have realized the limitations and constructed school climate instruments with more items and subscales in each dimension. The recent instruments also provide more psychometric evidence to support the reliability and validity of the instruments.

The empirical studies using the recent authoritative school climate instruments also found positive effects of support and structure on student outcomes. A series of studies conducted by Konold, Cornell and their colleagues found that middle and high school students in schools with high social support and discipline structure were more likely to have higher achievement, were more engaged in school activities (Cornell, Shukla, & Konold, 2015), and were less likely to drop out of school (Jia, Konold, & Cornell, 2015) and experience bullying victimization (Cornell et al., 2015). As previously mentioned, the DSCS is another instrument based on the authoritative discipline theory. Although there are few studies using the DSCS to examine the effects of school climate, the validation studies of the DSCS have showed a significant correlation between subscale scores and academic performance, suspension rates (Bear et al., 2011) and bullying victimization (Bear et al., 2014) at the school level.

2.3.2 Research Supporting the Subscales of the DSCS

Each subscale of the DSCS evaluates a specific aspect of school climate that has been linked to a range of student outcomes. The following paragraphs will elaborate and discuss the research findings supporting the subscales of the DSCS including teacher-student relationships, student-student relations, clarity of expectations, fairness of school rules, school safety, school-wide student engagement and bullying victimization.

Teacher-student relationships are a key construct of social support; a considerable amount of literature shows that students benefit from positive teacher-student relations. Student-reported teacher-student relations have been associated with academic achievement, academic competence (Fredricks, Blumenfeld, & Paris, 2004), peer acceptance (Hughes, Cavell, & Willson, 2001), school completion (Croninger &

Lee, 2001) and prosocial behavior (Pianta, Stuhlman, & Hamre, 2002; Wentzel, 2006). Teachers also benefit from supportive relations with students. Teachers who report more positive teacher-student relations are less likely to experience burnout and stress. These teachers also are more likely to be enthusiastic and motivated about their teaching practices (Grayson & Alvarez, 2008). Taken together, positive teacher-student relations contribute to both teachers and students' perception of a warm and caring learning environment.

Student-student relationships is another important subscale measuring the social support dimension of school climate. Ample research has shown that peer-support or student-student support plays a critical role in child development. Students' perceptions of peer support are positively associated with social-emotional adjustment (Brand et al., 2003), behavioral adjustment (Wang, Selman, Dishion, & Stormshak, 2010), achievement and academic initiative (Danielsen, Wiium, Wilhelmsen, & Wold, 2010). By contrast, peer-rejection is an influential risk factor predicting school avoidance (Buhs, Ladd, & Herald, 2006), problem behavior for boys (Nelson & Dishion, 2004), and poor achievement (Danielsen et al., 2010).

Respect for diversity is also a critical factor influencing students' and teachers' perceptions of school climate. Several studies have found that students, especially minority students, who experience more support and respect for diversity in schools are more likely to have higher academic achievement and report more favorable personal-social adjustment (Brand et al., 2003) and academic efficacy (Green, Adams, & Turner, 1988). In addition, Brand and his colleagues (2008) found that teachers' ratings of cultural pluralism are related to student achievement, as well as academic and behavioral adjustment. It is important that we gauge the respect for diversity in schools because doing so greatly contributes to evaluating school climate and understanding student achievement and adjustment.

It is generally accepted that clear and consistent behavioral expectations are the most effective strategy to manage students' behavioral problems in classrooms and handle discipline problems in schools (Bear, 2010b; Brophy, 1996). Several school-wide prevention programs promote positive school climate by emphasizing clear behavioral expectations and establishing fair school rules, such as School-Wide Positive Behavior Support (Bradshaw, Mitchell, & Leaf, 2010; Sugai & Horner, 2009). The studies evaluating the effectiveness of SWPBS have shown that the schools with high-level implementation are more likely to have clear and consistent behavior expectations in the classroom, hallway, and the whole school. As a result, suspension and discipline referral rates decrease and student achievement improve in these schools (Bradshaw et al., 2010). Clear and consistent behavioral expectations inform students of behavioral rules and establish organized classroom management systems.

A number of school climate instruments consider fairness of school rules an important factor in maintaining an organized classroom and dealing with discipline problems. For example, in the School Climate Survey (Zullig et al., 2014; Zullig et al., 2010), the Order and Discipline subscale assesses individuals' perceptions of the extent to which disciplinary policies are fair in classrooms and schools. The Inventory of School Climate (Brand et al., 2003; Brand, Felner, Seitsinger, Burns, & Bolton, 2008) has a subscale measuring students and teachers' perceptions of the extent to which rules and expectations are consistent and clear. Also, a number of studies have revealed that students are more likely to feel competent in academic activities and less likely to experience peer victimization in the schools with clear and fair rules (Brand et al., 2003; Gottfredson, D. C., Gottfredson, & Hybl, 1993; Gottfredson, G. D., Gottfredson, Payne, & Gottfredson, 2005).

The importance of student engagement has been emphasized by a number of school climate instruments in measuring a safe and positive school climate (e.g. Maryland Safe and Supportive Schools Climate Survey). Particularly, the aforementioned USDOE S3 model includes engagement as a main dimension of school climate instruments. The Student Engagement School-wide subscale measures the degree of attention and effort that most students show when they are learning in schools. This subscale is different from those instruments targeted at individual students' assessments of their own level of engagement; it is meant to assess a school-level phenomenon based on individuals' perceptions of the whole school as opposed to themselves. As noted in the landmark review of school engagement (Fredricks et al., 2004), "School engagement is seen as an antidote to such signs of student alienation." Several studies have found that student engagement is associated with academic achievement, adjustment, and school completion (Brand et al., 2008; Furlong et al., 2005).

A number of studies have found that students are less likely to experience bullying victimizations in schools with an authoritative climate (Berg & Cornell, 2016; Gerlinger & Wo, 2016). The research findings from the iconic Olweus Bullying Prevention program also characterized an authoritative school climate as the key to preventing students' aggressive behaviors (Olweus, 1994). Although these studies focused on bullying prevention, recent research has begun to address the importance of bullying in measuring school climate (Bandyopadhyay, Cornell, & Konold, 2009;

Klein, 2012). This consideration is reasonable because bullying victimization is closely related to school safety and student-student relations, which are two critical factors of school climate. Thus, the DSCS includes Bullying Victimizations School-wide as a subscale to measure school-level perceptions of bullying.

We can draw several conclusions from the theoretical frameworks of school climate, the development of various school climate instruments, and the research using and supporting the instruments:

1. School climate is a school-level phenomenon influenced by individuals' experiences and perspectives. It is best measured at both the school level and individual level.

2. It is preferable to evaluate school climate from all members of a school community in order to provide more comprehensive and unbiased information.

3. Since improving school climate is the focus of numerous bullying and discipline intervention programs, it is desirable to use overall scale and subscale scores predicting desired outcomes.

4. When school climate is used to evaluate the effectiveness of intervention programs, it may be measured longitudinally along with the implementation of such programs.

2.4 Methodological Issues in Measuring School Climate

The process of measuring school climate is complicated if the goal is to produce a comprehensive measure and use it as an indicator of school effectiveness.

Establishing a reliable and valid measure is crucial when measuring school climate. In addition, there are complicated methodological problems must be addressed, including reliability, measurement design and procedures, and scaling methods for school climate overall scores and subscale scores. G theory provides a convenient and efficient approach to evaluating the reliability of school climate instrument, optimizing measurement designs and procedures, and calculating overall school climate and subscale scores.

2.4.1 Demonstrating Reliability

A number of studies have discussed the misconception of reliability (Vacha-Haase, Henson, & Caruso, 2002), especially the misuse of internal consistency (i.e. Cronbach's α) as the only reliability coefficient that matters (Henson, 2001; Streiner, 2003). In the memoir by Cronbach (2004), he clarified none of the formal definitions of reliability treat internal consistency as reliability. There are different reliability coefficients that are critical to school climate research, such as test-retest reliability and interrater reliability. A test-retest reliability coefficient indicates the extent to which an instrument maintains a stable result over multiple measurement occasions. Interrater reliability refers to the degree to which raters' responses agree. Regarding school climate instruments, most studies only provide internal consistency coefficient of scale scores as an indicator of reliability (Bear et al., 2011; Bear et al., 2014; Bear, Yang, & Pasipanodya, 2015). Even though several studies have investigated school climate longitudinally and examined different raters' perceptions of school climate, none of these has reported test-retest reliability coefficients or interrater reliability coefficients. One possible reason is that calculation of such coefficients is quite complicated when a multi-level data structure must be taken into account. The other reason is that these two coefficients are commonly reported in the achievement testing literature rather in school climate research.

It is important to note these reliability coefficients are critical per se, but that they also have a particular meaning in school climate research. First, the analytic unit of school climate research is more focused on school-level climate than individual-level perceptions. The traditional methods of analyzing organizational level reliability and measurement error are inadequate (Cronbach, Lee J., Linn, Brennan, & Haertel, 1997). This has also been documented in the multilevel modeling literature through discussion of ecological fallacies, which refer to a generalization of organization-level effects to the individual-level or vice versa, and should be avoided (Robinson, 2009). In addition, investigation of reliability provides empirical evidence supporting the validity of instruments because conceptions of reliability and validity are closely intertwined (Cronbach, Lee J. et al., 1997). In order for an instrument to have high construct validity, the measurement structure must be stable. Most school climate instruments claim to have high construct validity; however, few studies have examined the stability of measurement structure of school climate at the school level (Brand et al., 2003; Brand et al., 2008).

Estimating reliability through traditional methods that derive from classical test theory is also limited because the methods estimate different reliability coefficients in separate calculations. In contrast, the prosed G theory has the advantage of analyzing different coefficients simultaneously and efficiently (Yin & Shavelson, 2008). In general, reliability coefficients can be considered a signal to noise ratio. For example, interrater reliability can be viewed as a function of the variance of between raters relative to error variance (i.e., “noise”). When measuring school climate, multiple factors contrinute to variance in measurement, such as rater groups, grade levels, and student body groups (i.e. schools). G theory has the flexibility and strength in identifying and distinguishing the “signal” and “noise” associated with different sources of variance, and it provides options to reduce contributions of noise to measurement results (Bloch & Norman, 2012).

2.4.2 The Need for Optimized Measurement Designs and Procedures

After identifying variations in measurement designs and procedures (e.g., number of items, raters, rater groups, etc.), a researcher may wonder whether it is possible to improve the reliability of measurement and how to obtain such optimizations. G theory can help researchers address the issues of improving reliability by changing measurement designs and procedures (Cronbach, Lee J. et al., 1997). For example, when measuring school climate: How many items and/or respondents are needed to obtain a reliable school climate score for a school? How many students, teachers, and parents are needed to evaluate school climate scale reliably? Answers to these questions can contribute to simplifying the complicated procedures involved in measuring school climate.

Decision studies (D studies) within G theory framework are intended to optimize measurement procedures. The role of a D study in a measurment design is similar to that of statistical power analysis in experimental design. Power analysis estimates the minimum sample size required to detect a real effect of a given value. In contrast, D studies are used to calculate how many items, raters, respondents, or occasions are needed to obtain a reliable measure, especially in performance assessment (Cronbach, Lee J. et al., 1997). It is worth mentioning that D studies can also be used in behavioral rating scales to optimize measurement procedures efficiently (Mashburn, Downer, Rivers, Brackett, & Martinez, 2014; Volpe, McConaughy, & Hintze, 2009). For example, Mashburn and his colleagues (2014) found that D studies can substantially reduce monetary cost of observation studies as well as improve the statistical power of an efficacy study using the optimized measure.

Although administration of a school climate survey can be much less expensive than an observation study, it still requires significant time and money. Survey administration is only one component of state-level prevention programs, but the administration or subsequent analysis may cost much more if done without cost-effective measurement procedures. For instance, during each of the past six years, the DSCS has been administered to approximately around 40,000 students, 15,000 teachers, and 5,000 parents in 140 schools each year. The cost of paper printing and survey scanning is significant. In addition, as increasing numbers of schools implement the DSCS and associated programs, the cost continues to rise. By applying G theory to the DSCS, it may be possible to mathematically determine the most cost-effective measurement procedures for achieving adequate reliability.

2.4.3 Scaling Issues in Measuring School Climate

Methods used to derive scores from school climate instruments are a critical component influencing the quality of information produced. The calculation of scale and subscale scores is generally determined by the format of measurement, the measurement model of an instrument, and the process of scale construction. Most school climate instruments use a Likert scale as the format of measurement. It assigns numbers to respondents' agreement levels for each statement or survey item. The summative assumption of a Likert-type scale allows researchers to calculate the scale scores by adding (or averaging) all the item scores. Except for the format of measurement, the calculation of subscale scores also depends on the measurement model of an instrument and scale construction. The measurement model provides theoretical justification for calculating subscale scores; the results from a factor-analytic model can help researchers determine empirically which items fall into which subscale. Without a theoretically sound and empirically validated measurement model, an instrument may not accurately assess the latent construct. The process of scale construction ensures the quality of an instrument and provides empirical justifications for calculating subscale scores. This process includes defining the concept and its construct, creating an item pool, determining the format for measurement, reviewing items, considering validation items, administering items to a development sample, evaluating items, and optimizing scale length (DeVellis, 2016).

After going through the procedure of scale construction, researchers can use various statistical methods to calculate scale and subscales scores (DeVellis, 2016). Non-refined methods and refined methods are two broad classes of these statistical approaches (DiStefano, Zhu, & Mindrila, 2009). Non-refined methods calculate scale scores by using summative procedures, which are relatively easy to compute and understand. The unit-weighting method is a frequently used non-refined method in which scale scores are calculated as a simple sum or average of item scores. Refined methods calculate standardized scale scores by using more complex mathematical transformations, such as multivariate scaling methods. Latent trait models and mathematical data reduction methods are two broad types of multivariate scaling methods. Both models aim at reducing a large amount of raw data into smaller numbers of variables. The main difference is that latent trait model takes measurement error into account while mathematical data reduction methods ignore measurement error. Furthermore, the applications of latent trait models usually involve latent variables (e.g. attitudes and feelings). School climate and its subscales are considered to be latent variables. However, researchers commonly use the unit-weighting sum or average scale scores to quantify school climate (Bear et al., 2011; Griffith, James, 2000).

Although using unit-weighting sum and/or average scale and subscales scores is typically acceptable for exploratory studies (Tabachnick, Fidell, & Osterlind, 2001), the limitations of using this method to quantify school climate may be particularly problematic in the following aspects: (a) differential item weights on scales and subscales, (b) nesting individual respondents within organizations, (c) handling data from multiple respondent groups, and (5) dealing with missing data at the item and respondent levels.

2.4.3.1 Differential item weights on scale and subscales

The unit-weighting method calculates the arithmetic sum or mean of items scores on a scale, ignoring item weights. As discussed, school climate is a latent construct that cannot be directly measured. The items of school climate instruments are considered to be observed consequences determined by individuals' perceptions of school climate and measurement error. The relationships between items and scales are not equal - "School rules are fair" is more relevant to subscale fairness of school rules than "The consequences of breaking the school rules are fair." This is a reality that the unit-weighting method disregards. The factor analytic model, on the other hand, one of the latent traits models, is often used to support the measurement model of school climate and scale items which indirectly measure school climate. In using this model to analyze data, the covariation among a set of items is examined to obtain information on their underlying latent variables, also called factors. The relationship between items and factors is regarded as factor loading. Different items may load different weights on the same factor. For instance, in Bear and his colleagues' paper (2011), the factor loading of the item "Teachers care about their students" on the teacher-student relations subscale was .53 and the factor loading of "Teachers treat students with respect" on the same factor was .38. The first item is weighted more heavily on this factor than the second. As such, the scale and subscales scores produced by factor analysis are adjusted by factor loadings and factor correlations, meaning that these scores contain the variations between items and factors. By contrast, the unit-weighting method constrains items to equal weights in calculating scale and subscales scores. As a result, unit-weighting may provide a lesser quality representation of latent constructs.

Another disadvantage of using the unit-weighting method is the disregard of measurement error. Measurement error represents the inherent variability between the measured value of an object and its true value. All measurement tools and procedures result in some degree of measurement error, which influences the precision of evaluating individual perception. For example, during the process of survey administration, individuals' moods may influence their responses, and alternative wording of individual items can produce changes in responses. Factor analytic models take measurement error into account when producing scale and subscales scores. However, the unit-weighting method considers each item response an observed variable, which may lead to biased estimation of scale and subscales scores. For example, some subscales are theoretically orthogonal (i.e. uncorrelated) but the unit-weighting subscale scores may be highly correlated (Brown, 2015; Grice, 2001).

These biased estimations risk the precision of scale and subscale scores and, in the worst cases, may lead to incorrect recommendations to educational practitioners.

Clearly, factor analytic models show more strengths than unit-weighting methods in scaling latent variables. In the field of school climate research, most studies use factor analytic models to support the measurement model of instruments but rarely used this approach for calculating factor scores. This may be due to the stage of development in methods for measuring school climate—most school climate studies are conducted as exploratory studies, which are less focused on psychometric measurement models, and more focused on basic validation of school climate surveys. For example, the DSCS has been developed and administered in Delaware for 12 years, and most published articles using the DSCS are validation studies of the survey.

In addition, with the advance of multilevel modeling techniques, the relative importance of the nested structure of school climate—the variability between perceptions at the level of the individual and at that of the whole school—has been the subject of considerable debate. Ignoring the nested structure risks biased statistical estimation and conceptual misunderstanding (Huang, J., 2012). More sophisticated factor analytic models, such as the multilevel factor analytic model, may be more appropriate and applicable to school climate scale and subscales scores. Due to the complexity of multilevel factor analytic models, only one notable school climate instrument, the Authoritative School Climate Survey, has been validated with that method (e.g. Konold et al., 2014).

2.4.3.2 Nesting of individual respondents within organizations

As discussed, multilevel scaling methods consider the nested nature of the data structure in measuring school climate. Within a school, students may have different perceptions of school climate due to their own experiences. Different schools are composed of different individuals and are influenced by different school districts. The DSCS is one of the few school climate instruments that has noted the importance of a nested data structure of school climate measurement. However, the current approach of validating the DSCS is simply to subtract the group mean from individuals' scores in running confirmatory factor analysis (CFA) (e.g., Bear et al., 2011). In other words, this approach only considers the variance within each school and discards the information about how school climate can be interpreted as a school-level phenomenon.

In addition, this group-centering approach may lead to complicated issues of scaling school climate scores when taking grade levels, respondent groups, and years into account. The current approach involves CFA for all grade levels and then tests measurement invariance across different grade levels. Even though this is a legitimate way to prove the factor structure and measurement model, there is no evidence showing the estimation of variance and covariance components among different schools nested in grade levels. Furthermore, the current approach analyzes the data for different rater groups separately rather than considering individuals nested within rater groups in a given school. Lastly, the DSCS has been administered in schools for over 10 years and it is unknown whether the factor structures have remained the same throughout that time.

2.4.3.3 Handling data from multiple respondent groups

In order to provide a more comprehensive and unbiased evaluation, it is preferable to measure school climate based on the perceptions of all members in a school community. As mentioned, the quality of being multi-informant is a unique strength of the DSCS. Few school climate instruments published in peer-reviewed journals assess students' and teachers' perceptions of school climate. The DSCS is the only school climate instrument that measures students', teachers' and parents' perceptions. However, the current scaling method of the DSCS uses a unit-weighting method to calculate the individual-level school climate scores separately for each respondent group. This simple average calculation potentially ignores the different importance of specific items and rater groups. It also forces analyses of school-level data to be conducted separately for each respondent group.

Fortunately, latent trait models such as CFA offer techniques to address item-weighting and combining information across multiple rater groups. As discussed, CFA can produce factor scores using factor loadings to determine item weights, also called the precision weighting method. Precision weighting is superior to unit-weighting when items contribute unequally to the reliability of scores (Peters, 2014). Similarly to CFA models, the proposed G theory models[1] can also produce precision-weighted "universe scores" that are analogous to factor scores. The G theory can also assign weights to items while considering all possible sources of measurement variance. In other words, G theory also determines the optimal weighting of respondent groups. Lastly, the G theory universe scores are empirical Bayes estimates, which increase reliability by "borrowing strength" from the full dataset (Raudenbush & Bryk, 2008).

[1] *Note.* A more detailed review will be provided in later sections.

2.4.3.4 Handling missing data at the item and respondent levels

Within school climate literature, current methodologies have two different approaches to handling missing data: (a) full information maximum likelihood (FIML) and (b) listwise deletion. The first method is usually used for handling missing data in CFA models. In previous validation studies, FIML was used to estimate model parameters with missing data because it has been shown to produce unbiased standard error and parameter estimates. The results of CFA using FIML to produce standardized factor scores can appropriately represent the latent structure of school climate, despite the missing data. However, the unbiased FIML factor scores have not been produced in previous CFA analyses or used for further descriptive and inferential analysis.

Listwise deletion is used for the calculation of unit-weighting scale scores when missing data occurs. Particularly, scale score is the average of raw score responses for each item on a scale after deleting the records with missing data. Unfortunately, listwise deletion omits any participants who did not answer one or more items on a scale when calculating scale scores. This approach is potentially biased because it may not preserve the important characteristics of the whole dataset, such as true correlations among scale scores and true mean values of scale scores.

Another criticism of listwise deletion is the loss of precision and power in statistical analysis with a smaller sample size. Small sample size is the most likely reason for a study being under-powered. In other words, it reduces the probability that a statistical procedure will detect a statistically significant difference, if such a difference truly exists in the population. In addition, listwise deletion combined with unit-weighting raises more concerns in longitudinal analysis because listwise deletion may lead to a much smaller sample size. For example, the DSCS has been administered to around 35,000 students in 145 schools since 2012. Since no information was collected that enables tracking individual students, the longitudinal analysis can only be conducted at the school-level. However, only 75% of schools administer the survey in any given year. As such, using listwise deletion when calculating scale scores for longitudinal analysis can reduce comparability of scales, decrease sample size and result in loss of statistical power.

By contrast, a latent trait model, such as CFA, has advantages in producing scale scores through model-estimation methods when dataset is incomplete. Besides the aforementioned FIML estimation, the proposed G theory model can also use Bayesian methods to impute missing data at the person and school levels. In addition, the G theory model treats changes in items throughout time as missing values (i.e., items that are modified are treated as new items) because it regards all of the items on a survey as randomly selected from a large item-pool. For instance, in comparison with the original DSCS (Bear et al., 2011), the recent survey has added a new item on the school safety subscale, "Students are safe in the hallway." Instead of deleting the new item in longitudinal analysis, G theory can approximate individuals' responses to this item based on their responses to other school safety items (i.e., imputation of responses) or it can calculate scale scores that reflect expected values had all questions been answered (i.e., full-information maximum likelihood). Some items of the DSCS have been changed or replaced. As a result, only 21 out of 35 items have been consistent across the past five years. Taken together, scale scores that are produced by G theory are stronger in longitudinal analysis than unit-weighting scale scores.

2.5 A Brief Review of Generalizability Theory

Because Generalizability theory (G theory) makes it possible to partition different sources of variance into separate components, it has become a valuable tool for studying reliability of behavioral observation data (Cardinet, Johnson, & Pini, 2010) and survey data (Schweig, 2014). G theory is defined as a methodological theory for (a) examining the reliability of the observable universe through the procedure of analysis of variance (ANOVA), and (b) optimizing the precision of measurement procedures (Cardinet et al., 2010). The G theory framework also provides computational formulas for reliability parameters and estimations of precision. In the sections that follow, I illustrate some of the key concepts underlying G theory.

2.5.1 Conceptualization of Reliability

To understand G theory, it is critical to understand first the concept of reliability. In both social and natural science fields, "error" is inherently associated with imprecision in measurement procedures (Cardinet, et al, 2010). In a physics experiment, the basic method to reduce random error is to measure an object for multiple times, using the mean of the observed values as the final quantity. Research in the social sciences experiences similar measurement error, but the most difficult obstacle is that the object of measurement is often hard to measure directly in the first place (e.g., a latent construct). The Classical Test Theory (CTT) is a fundamental psychometric theory to study measurement error and conceptualization of reliability. CTT regards the observed score (X) as a combination of true score (T, i.e., the latent trait) and random measurement error (ε). This fundamental formula for CTT can be expressed as:

$$X = T + \varepsilon \tag{1}$$

Relatedly, the conceptualization of reliability derived from CTT is expressed as a ratio of true score variance to observed score variance as follows:

$$\rho_{XX'} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_{T'}^2}{\sigma_{X'}^2} \tag{2}$$

If two parallel forms of a test exist (i.e., the original form is X, the alternative is X'), reliability is the correlation between their test scores. It also represents the variance ratio of true score to observed score based on either test (Allen & Yen, 2001). This formula indicates the extent to which true score variance is explained by the observed score. If the ratio is 1, it indicates the observed score of a test equals to the true score, and the test is constructed without measurement error. A $\rho_{XX'} = 1$ also implies that the two parallel forms of a test perfectly correlate with each other and the observed score of the original form equals the alternative form. Unfortunately, while this variance ratio is helpful for understanding the concept of reliability, this formula is not practical or operational for evaluating reliability because CTT assumes that true scores can never be directly measured, thus σ_T^2 is always unknown.

There are other interpretations of reliability which offer operational calculations of reliability coefficients, which are also based correlation paradigms. For example, internal consistency reliability, also known as Cronbach α, is the most widely used reliability coefficient (Webb, Shavelson, & Haertel, 2006). It measures all the possible alternative forms of the split-half reliability of a test and it represents the correlation between item scores on two halves of a test. In a more generalized situation, an item can be treated as a part of a test that can be divided into multiple pairs of items. Thus, coefficient α also represents the inter-item correlation on a test. The formula for coefficient is written (Cronbach, Lee J., 2004)

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum s_i^2}{s_t^2}\right) \tag{3}$$

In this formula, k is the total number of items on a test, s_t^2 represents the variance of total scores (t), and s_i^2 is the variance of items (i). If we visualize the formula in a two-way table (Table 2), item variance (s_i^2) is variance of total column scores and total variance (s_t^2) is based on the whole data set. Essentially, coefficient α reflects the consistency of items, which can be seen as the correlations between each column in Table 2 below.

Table 2.1: A Two-way Table of Person X Items in Cronbach's paper (2004)

Person × Item Score (X_{pi}) "Infinite" ("Population-Universe") Matrix *

	Item					
Person	*1*	2	…	*i*	…	$k \rightarrow \infty$
1	X_{11}	X_{12}	…	X_{1I}	…	X_{1k}
2	X_{21}	X_{22}	…	X_{2I}	…	X_{2k}
…	…	…	…	…	…	…
p	X_{p1}	X_{p2}	…	X_{pi}	…	X_{pk}
…	…	…	…	…	…	…
$n \rightarrow \infty$	X_{n1}	X_{n2}	…	X_{nI}	…	X_{nk}

Since computations of the coefficient can be carried out by most statistical software and the interpretation is comparatively simple, Cronbach's α has become the conventional coefficient in reporting reliability of test scores. However, it is rarely mentioned that Cronbach and his colleagues also proposed using a variance component approach to interpret internal consistency, which was influenced by the paradigm shift from correlation to analysis of variance in measurement theory. The variance components approach to α is a simplified case of generalizability theory (Cronbach, L. J., Gleser, Nanda, & Rajaratnam, 1972; Cronbach, Lee J., 2004). People are randomly chosen from a population and items on a test are randomly chosen from a large pool of items. This random sampling procedure of persons and items is illustrated in the infinite matrix (Table 2). The matrix also represents a data structure with a person cross item design (*p* x *i*) in a generalizability study (Cronbach, Lee J., 2004; Webb et al., 2006). The variance of observed total score ($\sigma^2_{X_{pi}}$) can be decomposed into person variance (σ^2_p), items variance (σ^2_i) and error variances (σ^2_{RES}):

$$\sigma^2_{X_{pi}} = \sigma^2_p + \sigma^2_i + \sigma^2_{RES} \tag{4}$$

According to the random effects ANOVA model (more detailed deductions in Webb, et al, 2006), Coefficient α can also be described as:

$$\alpha = \frac{\hat{\sigma}^2_p}{\hat{\sigma}^2_p + \frac{\hat{\sigma}^2_{RES}}{n_i}} \tag{5}$$

where $\hat{\sigma}^2_p$ and $\hat{\sigma}^2_{RES}$, are respectively the expected variance component of person and error. If extending the two-way matrix to more complicated cases, G theory can break down the total variances into systematic and random error variance. Simultaneously, it provides generalizability coefficients as indices of reliability and optimizations of measurement design based on estimated coefficients. What follows is a more detailed description of concepts and principles of G theory.

2.5.2 Key Concepts in G Theory

As previously explained, generalizability theory was born out of classical test theory and analysis of variance (Brennan, 2001), but it also has its own unique qualities. G theory has several strengths including 1) accounting for systematic error variances associated with measurement design and procedures (e.g. rater groups, occasions, and subject matters), random error variances, and interactions between sources of variation; 2) its ability to estimate random and fixed sources of variation; 3) the flexibility of evaluating dependability (an analog to reliability in CTT) of information on individuals' absolute and relative level of behavioral measurement (Cardinet et al., 2010; Yin & Shavelson, 2008). Before discussing how the DSCS can benefit from G theory, it is essential to describe and explain the concepts of G theory, including the universe of admissible generalization, the G (generalizability) study, and the D (decision) study (Brennan, 2001). These principles of G theory are also discussed relative to the specific context of measuring school climate using the DSCS.

2.5.2.1 Universe of Admissible Observations

In the framework of G theory, a universe of admissible observations contains all possible sources of variation that an investigator believes contribute to the measurement process, including the object of measurement and measurement facets. The *object of measurement* is the variability that an instrument intends to capture. School-level scale scores of school climate are the primary object of measurement because the DSCS aims to assess a school-level phenomenon. A *Facet,* which is analogous to a factor in ANOVA, represents a possible source of variation that would influence the accuracy of test (or scale) scores, such as rater, item, and occasion. Relating to the current study, there are three person-related facets. Since the primary object of measurement is school, individual respondents become a facet. They make up three *rater groups*, students, teachers, and parents, which occupy three different *grade levels*, elementary, middle, and high schools. In addition, there are two scale-related facets, including *items* and *subscales*. The last facet is *occasion*, indicating the time point that school climate has been evaluated. In sum, the universe of admissible observations consists of six facets including person, rater groups, grade levels, items, subscales, and occasions in the current study.

2.5.2.2 The Generalizability Study

A Generalizability study (G study) is an investigation process estimating the variability associated with all possible facets of measurement. After identifying the universe of observations, a researcher must consider the relationship between facets and facet-level sampling. The relationship between facets is either crossed or nested. If a facet is crossed with another facet, it means each level of the facet is associated with each level of another facet (Cardinet et al., 2010). For example, in this dissertation, each individual (p) answers every item (i) on the DSCS and this relationship between individuals and items is crossed, p x i. If a facet is nested within another facet, it indicates each level of the facet interacts with only one level of another facet. In this dissertation, a student is nested within the student rater group because a student respondent belongs to the student rater group and cannot be categorized into the teacher or parent rater group.

Facet-level sampling is defined as either fixed or random based on researchers' needs. A fixed facet suggests that all relevant levels of the facet are included and no sampling procedure is involved. An example would be rater groups in this dissertation. The rater-groups facet is fixed because it contains all three groups deemed relevant:

students, teachers, and parents. A random facet suggests random sampling in which the levels of a facet are randomly selected from a particular population. For example, individual respondents are randomly selected from the population at the local schools, and the items of the DSCS are randomly chosen from an infinite pool of items that could evaluate school climate. If measurement procedures include with both random and fixed facets, then the model of the G study is called a mixed design. In terms of the current study, a G study is carried out by estimating the variance components of school climate ratings through a mixed design with population (persons, *p*, nested in schools, *s*, which are nested in grade levels, *g*) crossed rater groups, *r*, and crossed with survey items nested in subscales (i: c) and crossed with measurement occasions (o). Using the notation of Brennan (2001), the model is represented as (*p: s: g*) x (*r)* x *(i: c*) x (*o*), where p, i, o, and s are random facets, and r, g, and c are fixed facets.

2.5.2.3 Optimization of Measurement Procedures (The Decision Study)

A G study aims to estimate the variance components for a universe of admissible observations. The goal of decision (D) studies is to use the estimation of the universe of admissible observations from a G Study to optimize measurement designs associated with specified universes of generalization (i.e., future planning of measurement procedures). It is important to note that a researcher can design different universes of generalization, which may include some or all the facets based on the result of the G study (Webb et al., 2006). Besides this attractive advantage, D studies also provide an unbiased universe score, which represents the expected score for a latent trait across all possible variations in measurement procedures (Brennan, 2001). The universe score is particularly useful for this study because it produces more accurate and precise scale scores for individuals and groups than unit-weighting methods.

There are two types of decisions can be made in D studies: relative and absolute decisions. As noted in the first treatment of G theory by Cronbach and his colleagues' (1972), *The Dependability of Behavioral Measurement*, G theory seeks to help when "an investigator asks about the precision or reliability of a measure because he/she wishes to generalize from the observations in hand to some class of observations to which it belongs." This generalization process is the core of G theory because it determines the estimations of parameters and interpretations of future sampling procedures in D studies.

Relative decisions are based on norm-referenced measurement error and focus on the ranked order of individuals (Brennan, 2001). For example, college admissions directors select candidates based on their abilities and performance relative to other candidates. Another example would be comparing school climate scale scores among schools within a school district. In relative decisions, relative error variance and generalizability coefficients are two commonly estimated parameters. Relative error variance $\sigma^2(\delta)$, also called measurement error for relative decisions, is the difference between an individual's observed deviation score and his or her universe deviation score. Conceptually, it is estimated as the sum of the variance of interactions between the object of measurement and facets in the measurement design. Relative error variance is an analogous to error variance in CTT, and the Generalizability coefficient is a similar concept to the reliability coefficient in CTT. Thus, the Generalizability coefficient ($E\rho^2$) is defined as the ratio of universe score variance $\sigma^2(\tau)$ to the sum of itself and relative error variance (Brennan, 2001). It can be written as:

$$E\rho^2 = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\delta)} \tag{6}$$

Absolute decisions relate to domain-referenced measurement error and emphasize the level of performance rather than ranking order, which is denoted $\sigma^2(\Delta)$. For example, to obtain SAS certification, one must achieve a minimum passing score of 70% on the examination, regardless of other examinees' performance. Relating to the current study, one possible absolute decision may be whether an individual feels sufficiently supported and safe in a school or whether parents in a school perceive a sufficiently warm and structural school climate. In abolute decisions, the two commonly estimated parameters are absolute error variance and index of dependability. Absolute error variance $\sigma^2(\Delta)$, also called measurement error for absolute decisions, is the difference between an individual's observed score and his or her universe score. Interpreting it from a variance decomposition view, absolute error variance is calculated as the variance of main effects for all facets and interactions between facets and the object of measurement (Brennan, 1983; Shavelson & Webb, 1991; cited in Lynch & McNamara, 1998). Absolute error variance is usually equal to or larger than relative error variance because it includes the variance of main effects for all facets, whereas the relative error variance does not.

The reliability coefficient related to absolute error variance is called index of dependability, which is denoted Φ. It is defined as the ratio of universe score variance to the sum of itself and absolute error variance:

$$\Phi = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\Delta)} \tag{7}$$

From equations (5), (6) and (7), we can conclude that the general formula of reliability coefficients is essentially a variance ratio of the variance accounted for the latent trait over the sum of itself and error variance. Another way of thinking about it is regarding the ratio of a signal to the sum of that signal and surrounding noise. In a particular D study, the G coefficient and index of dependability are computed based on sample sizes and the specification of measurement designs, such as the facet-level sampling and relationships between facets.

The final step of D studies is the process of efficiently and economically optimizing measurement designs. Broadly speaking, there are two commonly used approaches to accomplish this goal through a D study. The first approach is to manipulate the number of facet levels sampled. The number of levels for each source of variation influences the magnitude of error variance. This approach is similar to using a Spearman-Brown formula to decide how many items to include on a test (Shavelson & Webb, 1991). The second approach is to change the design of measurement, such as changing a fixed facet to random facet or changing the relationship between facets from crossed to nested. Through those approaches, a researcher may calculate how different designs can influence the generalizability coefficient and index of dependability.

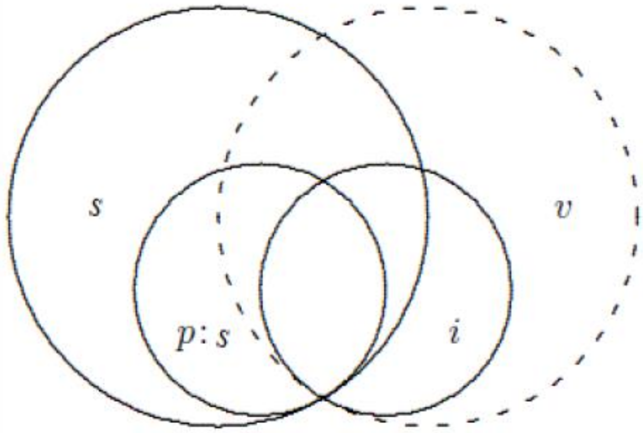
2.5.2.4 Need for multivariate generalizability theory

The previous sections on univariate generalizability theory provide foundations to understand the conceptual framework and analytic procedures of generalizability theory models. Multivariate generalizability theory is a big step forward as it allows parallels of universes and their composite in the admissible observations. By adding matrix computations to its algorithms, multivariate G theory has the capacity for analyzing variance and covariance of multiple universe scores of an instrument, especially for the instruments designed from a table of specifications (i.e., same as factor structure in scale development). For instance, the subscales of the DSCS can be treated as a linked facet (essentially a fixed facet) to connect each subscale to the total scale. As such, multivariate G theory is a more powerful analytic tool for analyzing the relative contributions of facets to the DSCS scale score than the univariate G theory model.

The characteristics of the DSCS measurement procedure can be summarized as a multi-dimensional, multi-level, multi-informant, and multi-year process. Theoretically, it is possible to include all possible facets in a multivariate G theory model, such as $(p{:}s{:}g)^{\bullet} \times r^{\bullet} \times i^{\circ} \times o^{\bullet}$, which means that all person in all schools and all grade levels answered all items across all rater groups and years. A superscript filled circle $^{\bullet}$ means that the facet is crossed with the linked facet (i.e. subscales) and an empty circle $^{\circ}$ means that the facet is nested within the linked facet. Different items are nested in each subscale of the DSCS. However, in practice, the survey administration of the DSCS did not allow such model tracking individual person across years. It is inevitable to break down the model by occasions (i.e. years). In addition, grade level differences may result in different variance structures (e.g., the reliability of middle school profile scores may not inform the sampling procedures in elementary or high schools), and the model should be examined separately by grade levels. The relationships between rater groups and person are confounded with the relationship between rater groups and schools. Each person can only be identified in one rater group and nested in a school, but a school can be rated by three rater groups.

If school profile scores are available, the rater group facet can be considered as a linked facet in a multivariate G theory design.

To summarize, by applying multivariate G theory to the DSCS, one possible approach is to break down the measurement procedures of the scale in a class-means design model, $(p^{\bullet} : s^{\bullet}) \times i$, with the DSCS subscales as the linked facet and a single-facet design model, $s^{\bullet} \times o^{\bullet}$ with rater groups as the linked facet. The representation of the class-means design $(p^{\bullet} : s^{\bullet}) \times i^{\circ}$ is shown in Figure 2.1.



$$\Sigma_s = \begin{bmatrix} X & X \\ X & X \end{bmatrix}$$

$$\Sigma_{p:s} = \begin{bmatrix} X & X \\ X & X \end{bmatrix}$$

$$\Sigma_i = \begin{bmatrix} X & \\ & X \end{bmatrix}$$

$$\Sigma_{si} = \begin{bmatrix} X & \\ & X \end{bmatrix}$$

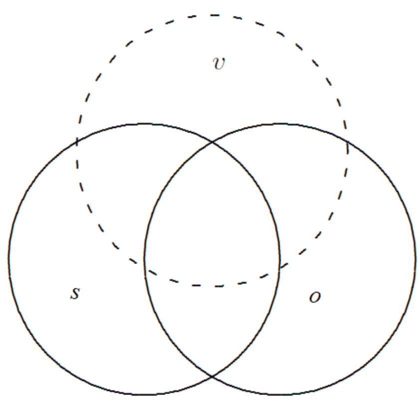
$$\Sigma_{pi:s} = \begin{bmatrix} X & \\ & X \end{bmatrix}$$

		v_1						v_2							
s	$p:s$	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9	i_{10}	i_{11}	i_{12}	i_{13}	i_{14}
1	1	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1	2	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1	3	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2	1	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2	2	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Figure 2.1: Representation of the *(p$^{\bullet}$: s$^{\bullet}$) x i°* design.

The dependent variable of this model is a person's rating on an item. This class-means design means that all person (*p*) in all schools (*s*) provided ratings of all items (*i*). Different items are associated with each subscale (*v*) of the DSCS. Again, a superscript filled circle $^{\bullet}$ means that the facet is crossed with the linked facet and an empty circle ° means that the facet is nested within the linked facet. As all persons in all schools are expected to answer the survey, the person facet and school facet are defined as two full matrices, $\widehat{\sum_{p:c}}$ and $\widehat{\sum_s}$ with variance components on the diagonal, covariance below the diagonal, and correlations above the diagonal. Other facets and the interactions between facets are defined as diagonal matrices with variance components on the diagonal. In this model, the object of measurement is school and D studies also focus on the aggregated unit, school. As a result, this model can provide a composite of multiple universe scores of the DSCS subscales for each school. The composite profile score for each school can be used as the dependent variable in the second single-facet deign.

The representation of the single-facet design model, $s^{\bullet}$ *x* $o^{\bullet}$, is shown in Figure 2.2.



$$\boldsymbol{\Sigma}_s = \begin{bmatrix} X & X \\ X & X \end{bmatrix}$$

$$\boldsymbol{\Sigma}_o = \begin{bmatrix} X & X \\ X & X \end{bmatrix}$$

$$\boldsymbol{\Sigma}_{so} = \begin{bmatrix} X & X \\ X & X \end{bmatrix}$$

	v_1							v_2						
s	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_1	o_2	o_3	o_4	o_5	o_6	o_7
1	X	X	X	X	X	X	X	X	X	X	X	X	X	X
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n_s	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Figure 2.2: Representation of the $s^{\bullet}$ *x* $o^{\bullet}$ design

This design represents that all schools (s) with all three rater groups (i) had school climate ratings across all the occasions (o). The model yields three full matrices, $\widehat{\sum_s}$ $\widehat{\sum_o}$ and $\widehat{\sum_{so}}$, with variance components on the diagonal, covariance components below the diagonal, and correlation above the diagonal. This model investigates the homogeneity between students, teachers, and parents' ratings of school-level school climate over the years.

Chapter 3

RESEARCH QUESTIONS AND METHODS

3.1 Research Questions

The primary research questions for this dissertation follow from the review of literature on theoretical frameworks of school climate instruments and methodological issues in measuring school climate described in the previous chapter. To summarize, prior literature on the theories and frameworks of school climate suggest that school climate is a multilevel phenomenon viewed by all members of a school community. Literature on measuring school climate suggests that its process involves complex measurement procedures, analysis of psychometric properties, and calculation of school climate scale scores. Following directly from these conclusions, the research questions guiding this dissertation are:

1. What does G theory tell us about the measurement design and precision of the Delaware School Climate Scale (DSCS)?
    a. What are the sources of variation (e.g. individuals within rater groups, schools, subscales, and the interactions among facets) contributing relatively more to school climate scale and subscale scores?
    b. Which types of raters' responses of school climate scale and subscale scores are more reliable across schools and occasions?
2. How can G theory optimize the efficiency and precision of the DSCS when simultaneously considering different sources of variance?

In the sections that follow, I described data sources, participants, measures, and data analytic procedures intended to address these research questions.

3.2 Data Sources

The research team at the University of Delaware Center for Disabilities Studies first developed the Delaware School Climate Scale (DSCS) to evaluate the behavioral intervention programs in the State of Delaware in 2006. The surveys also serve as brief and high-quality tools for schools to assess and improve various aspects of school climate. The research team has revised and updated the survey according to the needs of local schools, requests from local education agencies (LEA), and the results of psychometric analyses. In 2014, the state education agency was awarded the School Transformation Grant to assess and improve school climate, including implementation of a combined approach of School-wide Positive Behavior Support Intervention and Social Emotional Learning. The DSCS is the critical components of data integration to help LEAs examine and identify the strength and weakness of this combined approach and assess school climate. The surveys have three versions, including the student version (DSCS-S) for grades 3-12 (Bear et al., 2011), the teachers/staff (DSCS – T/S)(Bear et al., 2014) and home (DSCS-H) versions for all grades (K-12)(Bear et al., 2015).

In this dissertation, data collections of the DSCS in the past three years (2015-2017) were selected for the G study and D studies analysis. The DE-DOE sends out invitations to the DE-PBS schools to participate in administering the DSCS in the spring of each academic year. Usually, around 140 schools accept the invitations and assist data collections. In these schools, the building survey coordinators oversee the administrations of all three versions of the DSCS. The DE-PBS initiatives provided survey coordinators with detailed administration directions, such as format of survey (online or paper), assurance of confidentiality, methods for random sampling, and suggested sample size (more details are available on DE-PBS website). Teachers' surveys were completed online or paper surveys in November and December. Students and parents' surveys were completed by online or paper surveys in January and March of the academic year. Student completed the survey in schools and teachers were asked to ensure students answered all the items on the survey. Students brought the survey to home and parents/guardians completed the survey. Students brought the parents' surveys in sealed envelopes back to in school, which ensured that schools cannot have access to parents' responses. Online surveys data were collected using Qualtrics and paper surveys data were collected by Scranton machines. No identification information of individuals was collected but school identifications were available. The demographic variables were students' grade levels, ethnicity and gender based on students' and parents' responses.

3.3 Sample

According to the survey administration instructions, the DE-PBS initiatives aimed at sampling 100% of teachers and staff participations, 100% of students in elementary schools (3^{rd} grade to 5^{th} grade), 50% of students in middle and high schools with more than 600 students, 100% of students in school serving 300 or less students, 100% of parents in each school in the past three years.

The following sample selection procedures were used to choose the school, student, teacher and parent sample: (a) only regular public schools that participated the study were selected; (b) students responded "disagree" or "strongly disagree" to validation questions “I am telling the truth in this survey” and/or “I answered all items truthfully on this survey” were excluded; (c) staff members and administrators who were not also classroom teachers were excluded; (d) listwise deletion were used as mGENOVA does not allow missing response in the process of data analysis. After applying these procedures, the final sample was composed of total respondents including: 94,047 students, 11,004 teachers, and 38,719 parents in 149 schools from 2015 to 2017. It is likely that individuals completed the survey in more than one year. However, each survey response is counted as a distinct survey response since no identification information was collected that would allow linking of individuals’ responses over time.

In total, there were 41,754 elementary school students, 32,919 middle school students, and 19,374 high school students who responded the DSCS from 2015 to 2017. The percentages of student-reported ethnicities were 46.4% Caucasian, 24.2% African American, 3.7% Asian American, 12.6% Latino American, 10.6% multiracial, 0.3% Hawaiian, and 1.9% American Indian. Across all schools in the full sample, 48.7% students were boys. Parents’ responses were similar to students’. The percentages of parents-reported students’ ethnicities were 49.3% Caucasian, 19.8% African American, 6.1% Asian, 15.0% Hispanic, 10.3% multiracial, 0.1% Hawaiian, 0.5% American Indian. Among those students, 45.7% of theme were boys. The students and parent reported individual demographic information by year are presented in Table 3.1 and Table 3.2. Regarding teacher sample, the only available demographic variable is grade level. In the full sample, there were 5,542 teachers in elementary schools, 2,885 teachers in middle schools, and 2,577 teachers in high schools. The schools in this study accounted for roughly 75% of public schools in the state of Delaware. Although school-level demographic variables in the past three years were not available, state information may represent the variability in ethnicity. Across the state of Delaware, for the 2015-16 school year, the percentages of teachers in each ethnicity subgroup are 85.28% Caucasian, 10.29% African American, less than 1 % Asian American, 2.35% Latino American, less than 1% multiracial or other ethnicity.

Table 3.1: Student-reported students' demographic information

	2015		2016		2017	
	n	%	*n*	%	*n*	%
Grade Level						
Elementary	14066	45.8	15138	44.3	12550	43.1
Middle	10296	33.5	11594	33.9	11029	37.9
High	6369	20.7	7460	21.8	5545	19.0
Gender						
Boys	15051	49.0	16668	48.8	14108	48.5
Girls	15675	51.0	17509	51.2	15000	51.5
Race/Ethnicity						
American Indian	587	1.9	659	1.9	571	2.0
Asian	1139	3.7	1264	3.7	1026	3.5
African American	7563	24.8	8471	24.9	6654	22.9
Hawaiian	80	0.3	114	0.3	90	0.3
Hispanic/Latino	3859	12.6	4363	12.8	3589	12.4
Multi-Racial	3049	10.0	3719	10.9	3158	10.9
Caucasian	14244	46.7	15436	45.4	13931	48.0

Table 3.2: Parent-reported students' demographic information

	2015		2016		2017	
	n	%	*n*	%	*n*	%
Grade Level						
Elementary	10149	71.1	10481	73.7	6680	65.2
Middle	2958	20.7	2906	20.4	2777	27.1
High	1160	8.1	827	5.8	781	7.6
Gender						
Boys	6506	46.0	6383	45.3	4656	46.0
Girls	7651	54.0	7699	54.7	5472	54.0
Race/Ethnicity						
American Indian	58	0.4	80	0.6	60	0.6
Asian	821	5.9	890	6.5	570	5.7
African American	3028	21.9	2774	20.1	1678	16.7
Hawaiian	6	0.1	13	0.1	3	0.1
Hispanic/Latino	2050	14.8	2187	15.9	1422	14.1
Multi-Racial	1088	7.9	1349	9.8	1038	10.3
White	6773	49.0	6501	47.1	5294	52.6

3.4 Measure (Delaware School Climate Scale)

The technical properties of the Delaware School Climate Scales have been reported elsewhere (Bear et al., 2011; Bear et al., 2014; Bear et al., 2015; Bear et al., 2016). Standard psychometric analysis of the DSCS supported the construct validity of the measurement model, the internal consistency of the scale scores and the practical uses of the surveys. For each scale, confirmatory factor analyses were used to examine the structure of the measurement model and test measurement invariance across grade level, gender, and race. Cronbach α statistics were used to evaluate the internal consistency of the aggregated scale scores by school level. Concurrent validity of the surveys indicated the practical utilities of school level scale scores, which evaluated by the correlations between scale scores and academic achievement (i.e. percentages of students pass the state achievement tests). The number of items per subscale on the DSCS student, teacher, and home versions were presented in Table3.3.

Table 3.3: Number of items per subscale on DSCS-S/T/H

	N_i		
DSCS Subscales	Student	Teachers	Parents
Teacher-Student Relations	5	5	5
Student-Student Relations	5	5	5
Clarity of Expectations	4	4	4
Fairness of Rules	4	4	4
School Safety	3	3	3
Student Engagement-School-wide	6	6	-
School-wide Bullying	3	3	-
Teacher-Home Communications	-	4	4
Teacher-Staff Relations	-	4	-
Total scale	30	38	25

Students, teachers, and parents were asked in the survey to report "how much you think the following happens in this school" on a 4-point Likert scale, from 1 = Strongly Disagree to 4 = Strongly Agree within the following five common subscales:

Teacher-student relations. Five items measure respondents' perceptions of the extent to which teachers and staffs positively interact with students in the school, such as the respect and warmth toward students. The five items are: "Teachers care about their students." "Teachers listen to students when they have problems." "Adults who work at this school care about the students." "Teachers like their students." "Teachers treat students of all races with respect."

Student-student relations. Five items evaluate the extent to which respondents view caring and friendly interactions between students: "Students get along with one another." "Students are friendly with each other." "Students care about each other." "Students treat each other with respect." "Students respect others who are different." "Students respect others who are different."

Clarity of expectations. Four items assess the extent to which respondents perceive the behavioral expectations and rules are clear to understand and follow: "Rules are made clear to students." "Students know how they're expected to act." "Students know what the rules are." "This school makes it clear how students are expected to act." The structure of this subscale was quite stable, and no new item was added.

Fairness of rules. Four items measure the text to which respondents view the classroom rules, school rules, and the consequence of breaking those rules are fairly executed: "The school rules are fair." "The consequences of breaking the school rules are fair." "The school's Code of Conduct is fair." "Classroom rules are fair." A new item is added to the 2016 survey: "Adults in this school treat students fairly."

School safety. Three items measure the extent to which respondents generally feel about the level of being safe in their school: “Students are safe in the hallways.” “Students feel safe.” “Students know they are safe in this school.”

Also, students and teachers were asked to rate additional items on two subscales: school engagement and bullying school-wide. *School engagement* taps to how respondents generally feel students are engaged in school activities: “Most students try their best.” “Most students follow the rules.” “Most students turn in their homework on time.” “Most students work hard to get good grades.” “Most students feel happy.” “Most students like this school.” School-wide bullying includes three items that assess bullying victimization at the school level: “Students threaten and bully others." "Students worry about others bullying them." "Students bully one another."

In addition, teachers and were asked to rate four items on teacher-home communication subscale that evaluate the extent to which teachers communicate well with parents/guardians of students and showing respect: "Teachers listen to the concerns of parents." "Teachers do a good job communicating with parents." "Teachers show respect toward parents." "Teachers work closely with parents to help students when they have problems."

Teachers were also asked to rate five items on teacher-staff relations subscale that measures the quality of teachers, staffs, and administrators' interactions, as reflected in friendliness and collaborations among one another: "Teachers, staff, and administrators function as a good team" "There is good communication among teachers, staff, and administrators." "Teachers, staff, and administrators work well together." "Administrators and teachers support one another." "Teachers work well together in this school."

3.5 Analytic Plans

In situations involving psychometric data that is both nested (e.g., person within school) and crossed (e.g., respondent group, grade level), fully specified G theory models become very complicated. Ideally, there could be a multivariate G theory model (e.g., $(p{:}s{:}g)^{\bullet} \times r^{\bullet} \times i^{\circ} \times o^{\bullet}$) that specifies all conditions in the multiple universes of admissible observations including person, schools, items, respondent groups, grade levels, and occasions. However, it is not only mathematically complicated to calculate variance and covariance components with many fixed facets, but also conceptually confusing to visualize the relationship between facets. Persons are nested within schools that are crossed with respondent groups, but persons are also nested within respondent groups. It seems impossible to disentangle the confounding effects between person, school, and respondent groups in such a model, unless the ideal model is broken down into simpler models. To accomplish the goal of including all the facets, one possible approach is to use a simplified model to produce school profile scores, and then coordinate other facets into a second-order model based on school profile scores.

Another reason to break down the fully specified model is the limitation of data collections in longitudinal assessment. Although person and school are theoretically crossed with occasions, there was no individual identifier available to track respondents over the years. If occasions were added to a model, the school profile score would have to be linked for each school across all years. Furthermore, the theoretical model and variance structure may differ across grade level (e.g., the reliability of elementary school profile scores may not inform the sampling procedures in middle or high schools). If this occurs, then variance and covariance components will not be averaged across grade levels that show significant differences in variance components. Thus, grade level will not be included in the MGT models.

To summarize, the fully specified model will be broken down into two sets of multivariate G theory models, $(p^{\bullet} : s^{\bullet}) \times i^{\circ}$ and $s^{\bullet} \times o^{\bullet}$, with schools as the objects of measurement in both models. In the first set of analyses, a class-mean design, $(p{:}s)^{\bullet} \times i^{\circ}$, was estimated separately for each respondent groups ($n_r = 3$), grade level ($n_g = 3$), and occasions ($n_o = 3$); in total 3 x 3 x 3 = 27 models. The linked facet is subscales/factor in the measurement structure of the DSCS in the first model. All persons (*p*) in all schools (*p*) respond to all items (*i*). All persons and all schools contribute variability to all subscales, and different items are associated with each subscale. In each G study, seven matrices of variance and covariance were obtained, including two full matrices (i.e., variance and covariance matrices), $\widehat{\sum_{p:s}}$ and $\widehat{\sum_{s}}$, and three diagonal matrices (variance matrices), $\widehat{\sum_{\iota}}$, $\widehat{\sum_{s\iota}}$ and $\widehat{\sum_{p\iota:s}}$. D studies provide error variances, G and Phi coefficients of scale and subscales scores as well as school profile scores based on the variance and covariance matrices obtained in G studies. The school profile scores from the first stage of analyses were then used as the observed scores in the second order analyses to investigate the variability of occasions (i.e., years) on school climate scores reported by students, teachers, and parents in different grade levels. Due to the high volume of data, D study optimization procedures only focused on 2017 data because it is the most recent measurement situation to inform future sampling procedures. The optimization process illustrated by the D study shows the expected change in generalizability coefficients when manipulating the number of individuals within schools and the number of items in each survey.

The second order analyses, $s^{\bullet}$ x $o^{\bullet}$, were estimated separately for each grade level, in total three models. The linked facet is respondent groups. This model connects all possible facets while using school profile scores and maximizes the proficiency of multivariate G theory in analyzing the measurement procedures. The design also requires consistent assessment administration that each school asks all respondent groups to complete the survey every year. However, school districts or schools may withdraw, or they may include only one respondent group in the survey administration. As a result, 48 out of 78 elementary schools, 12 out of 28 middle schools, six out 24 high schools completed the survey every year from 2015 to 2017.

Those schools were selected in the second set of analysis, and the profile score of each school was considered as the dependent variable. Each G study provides estimations of three full matrices, $\widehat{\sum_s}$ $\widehat{\sum_o}$ and $\widehat{\sum_{so}}$. D studies produce estimations of error variance, G and Phi coefficients. The D study optimization procedures illustrated how coefficients change when manipulating the number of occasions.

All the G studies and D studies were analyzed using mGENOVA (Brennon, 2010), which was designed to perform multivariate generalizability theory analysis. Missing data is not allowed when analyzing variance components via mGENOVA. As such, listwise deletion, a method that includes only observations with complete responses, was used to handle missing data. The missing data rate at the individual level is reported in Table 1, ranging from zero for teacher responses in 2015 and 2016 to 24.4% for parent responses in 2015.

Table 3.4: Missing data rates

	2015	2016	2017
Elementary schools			
Students	10.7%	2.6%	1.0%
Teachers	0	0	0.9%
Parents	23.6%	11.8%	9.6%
Middle schools			
Students	7.9%	2.0%	0.4%
Teachers	0	0	0.2%
Parents	24.4%	11.6%	11.4%
High schools			
Students	7.8%	2.7%	4.2%
Teachers	0	0	1.4%
Parents	23.4%	9.7%	5.8%

Chapter 4

RESULTS

4.1 Results for RQ1

The first research question investigates what G theory tells us about the quality of measurement design and the precision of the DSCS scale and subscale scores. The results of G studies show what percentages of variance of scale and subscales scores that are attributable to "true" differences among persons, schools, respondent groups, items, grade levels, and occasions (i.e., years) in this phase. The analyses were conducted separately by grade level, respondent groups, and occasions, which in total, resulted in 27 multivariate class-mean models. Descriptive statistics of the DSCS scale and subscale scores were calculated using mGENOVA, and the results are presented in Table 4.1. As can be seen, students, teachers, and parents tended to have more favorable perceptions of school climate in elementary schools than middle and high schools. The raw average of the DSCS scale and subscale scores have increased from 2015 to 2017.

Table 4.1: Descriptive statistics of raw school climate scale and subscales scores

	Elementary		Middle		High	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
School climate						
Student						
2015	3.15	0.12	2.80	0.14	2.71	0.12
2016	3.17	0.13	2.83	0.14	2.74	0.09
2017	3.20	0.11	2.90	0.11	2.79	0.15
Teacher						
2015	3.24	0.19	2.96	0.21	2.92	0.18
2016	3.23	0.22	2.99	0.18	2.93	0.10
2017	3.26	0.17	3.00	0.19	2.99	0.15
Parent						
2015	3.36	0.09	3.09	0.13	2.73	0.23
2016	3.34	0.09	3.08	0.08	2.94	0.11
2017	3.37	0.08	3.14	0.10	3.14	0.19
Teacher-student relations						
Student						
2015	3.56	0.08	3.10	0.11	2.84	0.10
2016	3.56	0.10	3.09	0.13	2.84	0.07
2017	3.57	0.08	3.16	0.11	2.92	0.14
Teacher						
2015	3.51	0.14	3.32	0.11	3.27	0.12
2016	3.50	0.15	3.33	0.10	3.25	0.06
2017	3.53	0.10	3.34	0.11	3.30	0.10
Parent						
2015	3.42	0.08	3.15	0.11	2.81	0.24
2016	3.39	0.09	3.15	0.07	2.96	0.10
2017	3.42	0.07	3.19	0.08	3.18	0.19
Student-student relations						
Student						
2015	3.00	0.17	2.60	0.16	2.58	0.12
2016	2.99	0.19	2.61	0.16	2.59	0.11
2017	3.01	0.18	2.68	0.13	2.64	0.17
Teacher						
2015	3.12	0.22	2.79	0.25	2.83	0.17
2016	3.09	0.24	2.80	0.23	2.84	0.11
2017	3.12	0.21	2.85	0.23	2.90	0.16
Parent						
2015	3.16	0.11	2.82	0.16	2.63	0.28
2016	3.15	0.11	2.83	0.13	2.74	0.16
2017	3.21	0.10	2.92	0.14	2.97	0.17

Table 4.1(cont): Descriptive statistics of raw school climate scale and subscales scores

	Elementary		Middle		High	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Clarity of behavioral expectations						
Student						
2015	3.31	0.09	3.09	0.10	2.98	0.09
2016	3.29	0.12	3.10	0.09	2.99	0.07
2017	3.34	0.08	3.16	0.07	3.03	0.11
Teacher						
2015	3.40	0.20	3.17	0.19	3.04	0.21
2016	3.42	0.21	3.21	0.13	3.07	0.14
2017	3.44	0.14	3.22	0.18	3.12	0.18
Parent						
2015	3.43	0.08	3.24	0.11	2.59	0.23
2016	3.40	0.08	3.24	0.07	3.09	0.10
2017	3.43	0.07	3.27	0.08	3.26	0.21
Fairness of school rules						
Student	3.27	0.11	2.93	0.11	2.76	0.13
2015	3.22	0.14	2.93	0.12	2.75	0.11
2016	3.28	0.09	2.97	0.09	2.78	0.15
2017						
Teacher	3.33	0.20	3.17	0.25	3.06	0.17
2015	3.28	0.24	3.12	0.22	3.05	0.13
2016	3.33	0.16	3.15	0.17	3.12	0.14
2017						
Parent						
2015	3.39	0.08	3.16	0.11	2.85	0.20
2016	3.36	0.08	3.17	0.07	3.01	0.08
2017	3.38	0.08	3.20	0.09	3.18	0.22
School safety						
Student						
2015	3.22	0.13	2.78	0.21	2.74	0.20
2016	3.15	0.11	2.79	0.20	2.76	0.15
2017	3.26	0.13	2.91	0.16	2.85	0.20
Teacher						
2015	3.34	0.22	2.89	0.41	2.95	0.32
2016	3.28	0.25	2.94	0.28	2.97	0.19
2017	3.34	0.20	2.97	0.30	3.03	0.23
Parent						
2015	3.38	0.10	3.00	0.19	2.73	0.21
2016	3.34	0.10	2.97	0.16	2.87	0.15
2017	3.39	0.09	3.10	0.14	3.14	0.25

Table 4.1(cont): Descriptive statistics of raw school climate scale and subscales scores

	Elementary		Middle		High	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
School-wide engagement						
Student						
2015	3.31	0.09	3.09	0.10	2.98	0.09
2016	3.29	0.12	3.10	0.09	2.99	0.07
2017	3.34	0.08	3.16	0.07	3.03	0.11
Teacher						
2015	3.40	0.20	3.17	0.19	3.04	0.21
2016	3.42	0.21	3.21	0.13	3.07	0.14
2017	3.44	0.14	3.22	0.18	3.12	0.18
Bullying-victimization school-wide						
Student	2.56	0.26	2.43	0.22	2.54	0.18
2015	2.67	0.27	2.56	0.20	2.64	0.12
2016	2.78	0.26	2.62	0.17	2.68	0.19
2017						
Teacher	2.91	0.27	2.47	0.25	2.63	0.20
2015	2.95	0.29	2.56	0.20	2.72	0.13
2016	2.92	0.28	2.52	0.23	2.71	0.19
2017	2.56	0.26	2.43	0.22	2.54	0.18
Teacher-home communications						
Teacher	3.38	0.17	3.21	0.10	3.11	0.14
2015	3.37	0.15	3.21	0.09	3.06	0.06
2016	3.40	0.13	3.23	0.10	3.12	0.10
2017						
Parent						
2015	3.39	0.08	3.13	0.12	2.77	0.22
2016	3.38	0.08	3.13	0.07	2.94	0.08
2017	3.40	0.07	3.16	0.09	3.13	0.18
Teacher-staff relations						
Teacher						
2015	3.07	0.34	2.88	0.30	2.78	0.30
2016	3.08	0.36	2.92	0.34	2.80	0.28
2017	3.15	0.29	2.92	0.36	2.86	0.29

4.1.1 Results of G Studies in Elementary Schools

Table 4.2 to 4.10 show the variance and covariance of facets, the covariance between subscales at the person and school level, and the correlations between subscales at the school level in elementary school student, teacher, and parent groups from 2015 to 2017 sample. Figures 4.1 to 4.3 show the proportion of each variance component to the total variance. In these tables, $\widehat{\sum_s}$ represents the full matrix with school variance of each subscale on the diagonal, and covariances between subscales off the diagonal, correlations between subscales above the diagonal. $\widehat{\sum_{p:c}}$ is a half matrix within person within school variance of each subscale on the diagonal and covariance between subscales off the diagonal. $\widehat{\sum_\iota}$ represents a diagonal matrix with item variance of each subscale on the diagonal. Due to the limited space, each element of the diagonal matrix is presented in a horizontal direction. $\widehat{\sum_{s\iota}}$ represents a diagonal matrix with the variances of the interaction between items and schools of each subscale on the diagonal. $\widehat{\sum_{p\iota:s}}$ represents a diagonal matrix with residual variance of each subscale on the diagonal.

As shown in Figures 4.1 to 4.3, the results of the G studies were quite stable for elementary school students from 2015 to 2017. The greatest percentage of variance was attributable to the residual variances (over 50% of the total variance) in six out of seven subscales: teacher-student relations, school-wide engagement, clarity of behavioral expectations, fairness of school rules, school safety, and school-wide bullying. This indicates that there was substantial variability in elementary school students' ratings of school climate attributable to interactions between students, schools, items within each subscale, and other unknown sources of variance. The second largest variance component was the student within school variance (ranging from 31% in school-wide bullying in 2015 to 49.46 % in student-student relations in 2015), which indicates the variability attributed to students, and the interaction between students and schools. The school facet explained a relatively small variance (ranging from 1.47% in teacher-student relations in 2017 to 6.21% in school-wide bullying in 2017), which indicates the difference in school climate subscale scores varied to some degree between schools. The item facet also accounted for a small proportion of the total variance in each subscale (ranging from 1.40% in teacher-student relations in 2016 to 9.53% in school-wide bullying in 2015), which indicates low variability in item differences averaging across students and schools within each subscale. The interaction between school and item explained the least amount of variance (ranging from 0.16% in student-student relations 2015 to 1.03% in school-wide bullying in 2016). This indicates that a negligible amount of variability between schools is associated with their relative standing across individual items. Table 4.2 to 4.4 also present the estimated G studies covariance components at the student and school level and correlations for elementary schools across seven from 2015 to 2017. Universe score correlations among the school subscale scores were relatively high ranging from .48 (school-wide bullying and teacher-student relations in 2015) to .96 (student-student relations and fairness of school rules in 2016), indicating that the seven subscales are moderately to strongly related.

Table 4.2: Estimated variance and covariance component matrices for elementary school students in 2015

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULLY
$\hat{\sum}_{\boldsymbol{s}}$							
	0.007	0.738	0.756	0.873	0.819	0.809	0.484
	0.010	0.029	0.916	0.708	0.760	0.906	0.815
	0.007	0.018	0.013	0.670	0.875	0.834	0.652
	0.006	0.011	0.007	0.008	0.802	0.882	0.543
	0.008	0.014	0.011	0.008	0.012	0.874	0.564
	0.009	0.020	0.013	0.010	0.013	0.017	0.732
	0.010	0.036	0.020	0.012	0.016	0.025	0.067
	1.5%	4.6%	2.4%	1.4%	1.8%	2.9%	5.4%
$\hat{\sum}_{\boldsymbol{p}:\boldsymbol{s}}$							
	0.185						
	0.134	0.311					
	0.117	0.192	0.171				
	0.117	0.150	0.126	0.199			
	0.157	0.138	0.131	0.146	0.229		
	0.135	0.190	0.146	0.142	0.148	0.227	
	0.051	0.143	0.069	0.059	0.058	0.121	0.386
	42.3%	49.4%	31.0%	37.8%	34.9%	39.3%	31.0%
$\hat{\sum}_{\boldsymbol{i}}$							
	0.006	0.003	0.016	0.010	0.028	0.027	0.119
	1.4%	0.5%	2.8%	1.8%	4.2%	4.7%	9.5%
$\hat{\sum}_{\boldsymbol{si}}$							
	0.001	0.001	0.004	0.001	0.006	0.005	0.008
	0.3%	0.1%	0.7%	0.2%	0.8%	0.8%	0.6%
$\hat{\sum}_{\boldsymbol{pi}:\boldsymbol{s}}$							
	0.237	0.284	0.347	0.308	0.380	0.301	0.666
	54.2%	45.2%	62.8%	58.5%	58.0%	52.0%	53.4%
G	0.778	0.909	0.837	0.767	0.760	0.807	0.778
Phi	0.679	0.890	0.699	0.623	0.527	0.567	0.679
$\sigma^2(\delta)$	0.002	0.003	0.003	0.002	0.004	0.004	0.002
$\sigma^2(\Delta)$	0.003	0.004	0.006	0.005	0.011	0.013	0.003

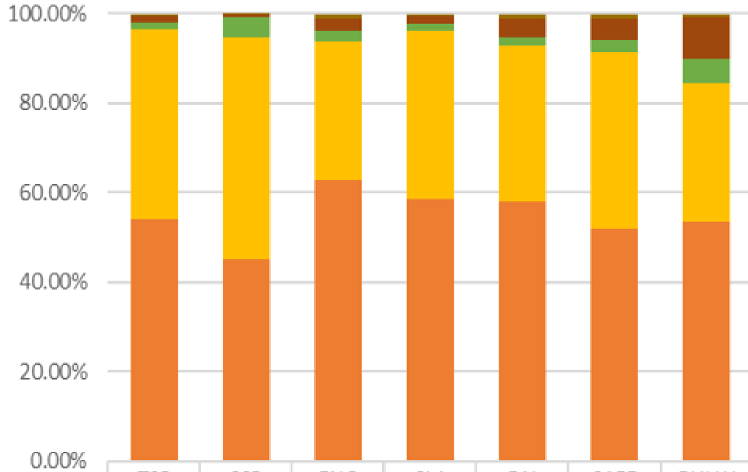
Note. TSR= teacher-student relations, SSR= student-student relations, ENG = school-wide engagement, CLA= clarity of behavioral expectations, FAI = fairness of school rules, SAFE= school safety, BULLY= school-wide bullying.

Table 4.3: Estimated Variance and covariance component matrices for elementary school students in 2016

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULLY
$\hat{\sum} s$							
	0.010	0.833	0.953	0.858	0.889	0.913	0.646
	0.016	0.036	0.824	0.749	0.960	0.941	0.853
	0.011	0.018	0.014	0.937	0.930	0.851	0.683
	0.010	0.017	0.013	0.014	0.858	0.860	0.646
	0.013	0.026	0.015	0.014	0.020	0.919	0.807
	0.010	0.019	0.011	0.011	0.014	0.011	0.693
	0.017	0.043	0.021	0.020	0.030	0.019	0.071
	2.3%	5.5%	2.5%	2.1%	3.3%	2.0%	5.9%
$\hat{\sum} p:s$							
	0.185						
	0.128	0.305					
	0.113	0.147	0.190				
	0.148	0.127	0.140	0.208			
	0.126	0.188	0.137	0.138	0.211		
	0.108	0.182	0.116	0.113	0.135	0.157	
	0.050	0.155	0.066	0.054	0.122	0.067	0.403
	42.2%	47.2%	35.4%	32.7%	35.3%	28.9%	33.8%
$\hat{\sum} i$							
	0.006	0.004	0.011	0.030	0.036	0.029	0.049
	1.4%	0.6%	2.0%	4.6%	5.9%	5.2%	4.1%
$\hat{\sum} si$							
	0.001	0.001	0.002	0.006	0.005	0.004	0.012
	0.2%	0.1%	0.3%	0.9%	0.7%	0.8%	1.0%
$\hat{\sum} pi:s$							
	0.236	0.299	0.319	0.377	0.325	0.341	0.655
	53.7%	46.3%	59.6%	59.4%	54.5%	62.8%	55.0%
G	0.862	0.936	0.869	0.798	0.847	0.845	0.915
Phi	0.782	0.915	0.742	0.558	0.560	0.621	0.790
$\sigma^2(\delta)$	0.002	0.002	0.002	0.003	0.004	0.002	0.007
$\sigma^2(\Delta)$	0.003	0.003	0.005	0.011	0.015	0.007	0.019

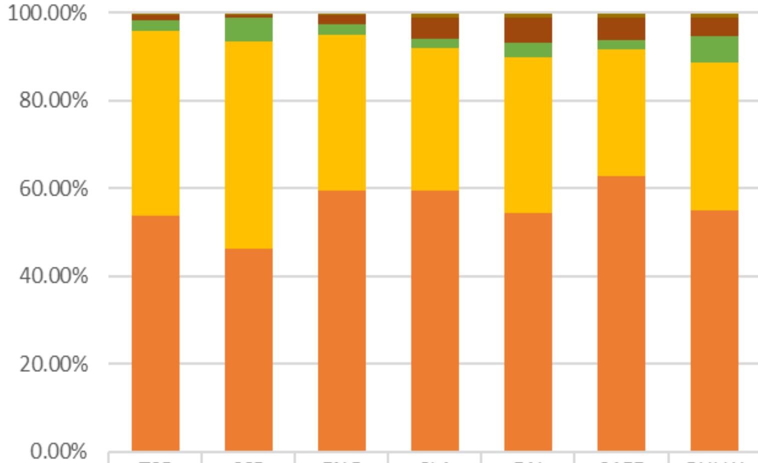
Table 4.4: Estimated Variance and covariance component matrices for elementary school students in 2017

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULLY
$\hat{\sum}_{\boldsymbol{s}}$							
	0.006	0.832	0.834	0.891	0.819	0.942	0.738
	0.012	0.031	0.935	0.678	0.805	0.925	0.920
	0.006	0.016	0.009	0.649	0.888	0.815	0.750
	0.006	0.010	0.005	0.007	0.758	0.904	0.598
	0.006	0.013	0.008	0.005	0.008	0.760	0.675
	0.010	0.022	0.011	0.010	0.009	0.018	0.859
	0.015	0.042	0.019	0.013	0.016	0.030	0.068
	1.4%	5.0%	1.7%	1.3%	1.2%	3.2%	6.2%
$\hat{\sum}_{\boldsymbol{p:s}}$							
	0.183						
	0.127	0.281					
	0.105	0.172	0.152				
	0.107	0.129	0.109	0.173			
	0.145	0.131	0.114	0.137	0.216		
	0.127	0.172	0.126	0.123	0.137	0.190	
	0.064	0.174	0.074	0.065	0.064	0.127	0.395
	43.0%	45.9%	28.7%	35.4%	35.0%	33.7%	36.2%
$\hat{\sum}_{\boldsymbol{i}}$							
	0.007	0.006	0.029	0.009	0.025	0.041	0.073
	1.5%	1.0%	5.5%	1.7%	4.0%	7.2%	6.7%
$\hat{\sum}_{\boldsymbol{si}}$							
	0.001	0.001	0.004	0.002	0.004	0.005	0.008
	0.3%	0.2%	0.7%	0.3%	0.6%	0.9%	0.6%
$\hat{\sum}_{\boldsymbol{pi:s}}$							
	0.227	0.292	0.332	0.298	0.362	0.309	0.545
%	53.5%	47.7%	63.0%	61.1%	58.9%	54.8%	50.1%
G	0.796	0.935	0.834	0.786	0.741	0.842	0.796
Phi	0.682	0.901	0.582	0.625	0.467	0.516	0.682
$\sigma^2(\delta)$	0.002	0.002	0.002	0.002	0.003	0.003	0.002
$\sigma^2(\Delta)$	0.003	0.003	0.007	0.004	0.009	0.017	0.003



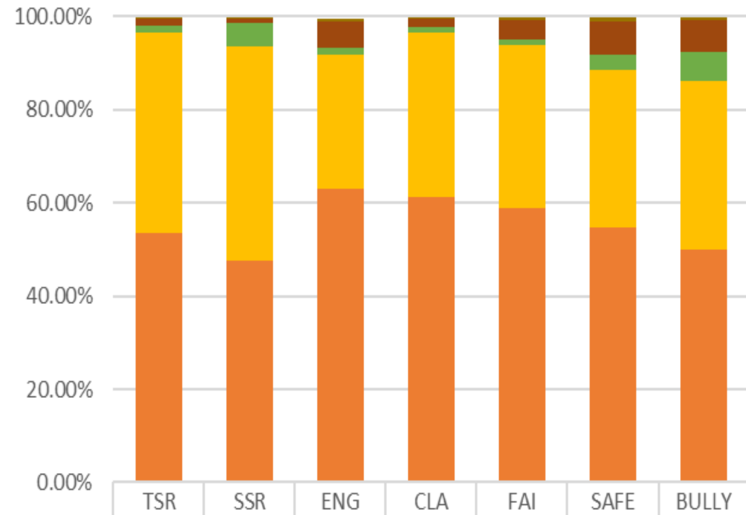
	TSR	SSR	ENG	CLA	FAI	SAFE	BULLY
si	0.30%	0.10%	0.70%	0.20%	0.80%	0.80%	0.60%
i	1.40%	0.50%	2.80%	1.80%	4.20%	4.70%	9.50%
s	1.50%	4.60%	2.40%	1.40%	1.80%	2.90%	5.40%
p:s	42.30%	49.40%	31.00%	37.80%	34.90%	39.30%	31.00%
pi:s	54.20%	45.20%	62.80%	58.50%	58.00%	52.00%	53.40%

Figure 4.1: Proportions of each variance component to the total variance for elementary school students in 2015.



	TSR	SSR	ENG	CLA	FAI	SAFE	BULLY
si	0.20%	0.10%	0.30%	0.90%	0.70%	0.80%	1.00%
i	1.40%	0.60%	2.00%	4.60%	5.90%	5.20%	4.10%
s	2.30%	5.50%	2.50%	2.10%	3.30%	2.00%	5.90%
p:s	42.20%	47.20%	35.40%	32.70%	35.30%	28.90%	33.80%
pi:s	53.70%	46.30%	59.60%	59.40%	54.50%	62.80%	55.00%

Figure 4.2: Proportions of each variance component to the total variance for elementary school students in 2016.



	TSR	SSR	ENG	CLA	FAI	SAFE	BULLY
si	0.30%	0.20%	0.70%	0.30%	0.60%	0.90%	0.60%
i	1.50%	1.00%	5.50%	1.70%	4.00%	7.20%	6.70%
s	1.40%	5.00%	1.70%	1.30%	1.20%	3.20%	6.20%
p:s	43.00%	45.90%	28.70%	35.40%	35.00%	33.70%	36.20%
pi:s	53.50%	47.70%	63.00%	61.10%	58.90%	54.80%	50.10%

Figure 4.3: Proportions of each variance component to the total variance for elementary school students in 2017.

For the elementary school teacher sample, the results of the G studies were quite stable from 2015 to 2017, as shown from Figure 4.4 to Table 4.6. The teacher within school facet explained the largest proportion of the total variance in seven out of nine subscales including teacher-student relations, student-student relations, clarity of expectations, school-wide bullying, school-home communication, teacher-staff relations, and school safety. Residual variance accounted for the largest proportion of the total variance in school-wide engagement and fairness of school rules subscales. Residual variance accounted for the second largest proportion of the total variance in the rest of the seven subscales. The school facet was the third largest variance component (ranging from 3.76% in teacher-student relations in 2017 to 22.34% in teacher-staff relations in 2016). The item facet accounted for a certain proportion of the total variance in each subscale, ranging from 0.37% in school safety in 2017 to 10.84% in fairness of school rules in 2015. The interaction between item and school facet explained a negligible proportion of the total variance ranging from 0.12% in student-student relations in 2017 to 4.15% in fairness of school rules in 2015. Tables 4.5 to 4.7 show estimated G-student covariance components at the teacher and school level and correlations for schools across seven subscales in elementary schools from 2015 to 2017. Universe score correlations among the school-level subscale scores were from moderate to relative high, which ranged from .357 (school-wide engagement and teacher-staff relations in 2017) to 1.00 (student-student relations and school-wide engagement in 2015). This indicates that the nine subscales are closely related. The almost perfect correlations between student-student relations and school-wide engagement suggest that the two subscales might measure redundant concepts based on teachers' perspectives.

Table 4.5: Estimated Variance and covariance component matrices for elementary school teachers in 2015

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULL	COM	STF
$\hat{\sum}s$									
	0.018	0.860	0.858	0.857	0.872	0.905	0.853	0.974	0.568
	0.026	0.049	1.007	0.765	0.814	0.945	0.994	0.809	0.400
	0.024	0.045	0.041	0.778	0.818	0.927	0.985	0.807	0.434
	0.023	0.034	0.031	0.039	0.915	0.884	0.788	0.839	0.679
	0.023	0.036	0.033	0.036	0.039	0.942	0.837	0.810	0.686
	0.027	0.046	0.041	0.038	0.040	0.047	0.972	0.863	0.638
	0.032	0.061	0.055	0.043	0.045	0.058	0.076	0.802	0.458
	0.023	0.031	0.028	0.029	0.028	0.033	0.038	0.030	0.510
	0.026	0.031	0.030	0.046	0.047	0.048	0.043	0.031	0.119
%	6.3%	17.8%	13.0%	11.1%	8.3%	15.0%	16.1%	10.3%	20.2%
$\hat{\sum}p:s$									
	0.123								
	0.077	0.133							
	0.075	0.102	0.102						
	0.106	0.088	0.090	0.199					
	0.115	0.086	0.090	0.157	0.172				
	0.102	0.101	0.095	0.124	0.125	0.152			
	0.054	0.087	0.072	0.064	0.072	0.097	0.214		
	0.116	0.077	0.081	0.105	0.110	0.103	0.060	0.168	
	0.111	0.093	0.091	0.142	0.156	0.122	0.079	0.130	0.372
%	42.8%	48.2%	32.1%	56.3%	36.8%	48.1%	45.4%	57.6%	63.2%
$\hat{\sum}i$									
	0.017	0.001	0.016	0.002	0.051	0.000	0.011	0.005	0.002
%	6.0%	0.3%	5.2%	0.5%	10.8%	0.1%	2.3%	1.9%	0.4%
$\hat{\sum}si$									
	0.003	0.002	0.009	0.002	0.019	0.004	0.002	0.002	0.001
%	1.0%	0.7%	2.7%	0.7%	4.2%	1.2%	0.4%	0.8%	0.2%
$\hat{\sum}pi{:}s$									
	0.126	0.091	0.149	0.111	0.186	0.113	0.168	0.086	0.094
%	43.9%	32.9%	47.0%	31.4%	39.9%	35.6%	35.8%	29.5%	15.9%
G	0.702	0.865	0.845	0.772	0.715	0.820	0.855	0.755	0.860
Phi	0.619	0.862	0.801	0.765	0.580	0.819	0.830	0.730	0.856
$\sigma^2(\delta)$	0.008	0.008	0.008	0.012	0.015	0.010	0.013	0.010	0.019
$\sigma^2(\Delta)$	0.011	0.008	0.010	0.012	0.028	0.010	0.016	0.011	0.020

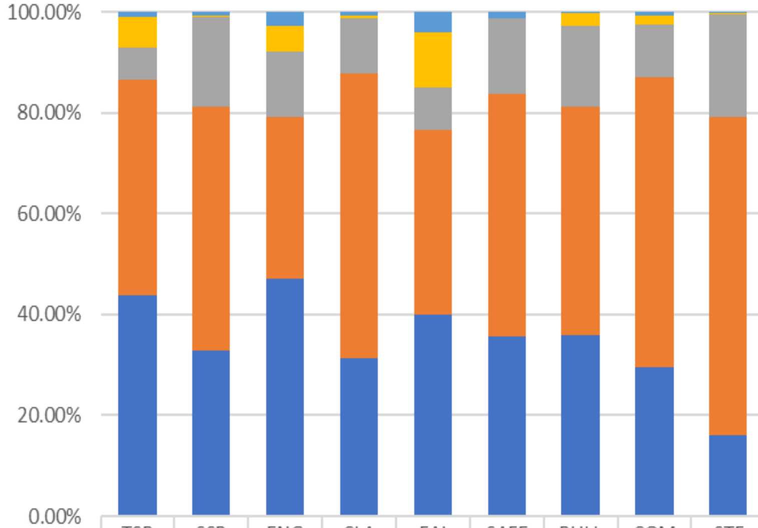
Note. TSR= teacher-student relations, SSR= student-student relations, ENG = school-wide engagement, CLA= clarity of behavioral expectations, FAI = fairness of school rules, SAFE= school safety, BULL= school wide bullying, COM= teacher-home communication, STF= staff-teacher relations.

Table 4.6: Estimated Variance and covariance component matrices for elementary school teachers in 2016

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULL	COM	STF
$\hat{\sum} \boldsymbol{s}$									
	0.021	0.937	0.922	0.848	0.889	0.925	0.912	0.972	0.822
	0.033	0.058	0.996	0.806	0.833	0.945	0.966	0.912	0.700
	0.031	0.056	0.055	0.812	0.831	0.947	0.959	0.946	0.688
	0.026	0.041	0.040	0.046	0.913	0.924	0.811	0.921	0.893
	0.032	0.049	0.047	0.048	0.059	0.934	0.848	0.848	0.805
	0.034	0.058	0.056	0.050	0.058	0.064	0.964	0.935	0.787
	0.038	0.066	0.064	0.049	0.059	0.070	0.081	0.917	0.712
	0.021	0.032	0.033	0.029	0.031	0.035	0.039	0.022	0.800
	0.043	0.061	0.058	0.069	0.071	0.072	0.074	0.043	0.131
%	7.6%	19.6%	16.5%	13.1%	12.0%	18.9%	16.5%	7.9%	22.3%
$\hat{\sum} \boldsymbol{p:s}$									
	0.134								
	0.067	0.140							
	0.071	0.111	0.107						
	0.110	0.089	0.089	0.192					
	0.111	0.091	0.095	0.156	0.181				
	0.096	0.114	0.106	0.131	0.133	0.157			
	0.056	0.105	0.088	0.083	0.093	0.115	0.212		
	0.118	0.076	0.075	0.108	0.110	0.093	0.065	0.160	
	0.113	0.107	0.111	0.144	0.171	0.151	0.108	0.140	0.363
%	48.3%	47.1%	32.4%	55.2%	36.6%	46.3%	43.2%	57.9%	61.7%
$\hat{\sum} \boldsymbol{i}$									
	0.010	0.002	0.014	0.003	0.043	0.000	0.024	0.004	0.002
%	3.7%	0.6%	4.3%	0.9%	8.8%	0.1%	4.9%	1.5%	0.4%
$\hat{\sum} \boldsymbol{si}$									
	0.001	0.001	0.002	0.002	0.020	0.004	0.001	0.001	0.001
%	0.5%	0.3%	0.7%	0.4%	4.0%	1.1%	0.3%	0.4%	0.2%
$\hat{\sum} \boldsymbol{pi:s}$									
	0.111	0.096	0.152	0.106	0.191	0.115	0.173	0.089	0.090
%	39.9%	32.3%	46.0%	30.4%	38.6%	33.8%	35.1%	32.3%	15.4%
G	0.718	0.875	0.883	0.797	0.781	0.851	0.858	0.693	0.867
Phi	0.671	0.870	0.850	0.787	0.684	0.850	0.807	0.671	0.865
$\sigma^2(\delta)$	0.008	0.008	0.007	0.012	0.017	0.011	0.013	0.010	0.020
$\sigma^2(\Delta)$	0.010	0.009	0.010	0.012	0.028	0.011	0.019	0.011	0.021

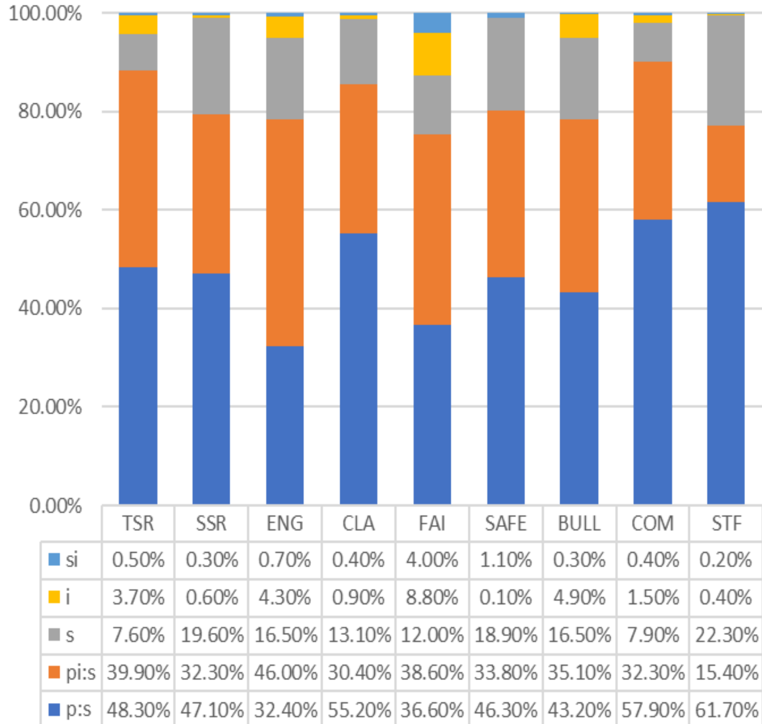
Table 4.7: Estimated Variance and covariance component matrices for elementary school teacher in 2017

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULL	COM	STF
$\hat{\sum} s$									
	0.010	0.852	0.908	0.772	0.925	0.892	0.800	0.971	0.583
	0.018	0.042	0.990	0.763	0.757	0.966	0.943	0.765	0.389
	0.020	0.043	0.045	0.759	0.775	0.966	0.937	0.852	0.357
	0.011	0.021	0.022	0.018	0.879	0.859	0.796	0.796	0.619
	0.015	0.025	0.027	0.019	0.027	0.880	0.748	0.894	0.626
	0.018	0.039	0.041	0.023	0.029	0.039	0.958	0.848	0.505
	0.023	0.055	0.056	0.030	0.035	0.054	0.080	0.742	0.384
	0.013	0.020	0.023	0.014	0.019	0.022	0.027	0.016	0.605
	0.017	0.023	0.022	0.024	0.030	0.029	0.031	0.022	0.084
%	3.7%	15.7%	13.8%	5.8%	6.3%	12.9%	16.6%	5.9%	15.9%
$\hat{\sum} p:s$									
	0.135								
	0.079	0.135							
	0.079	0.108	0.105						
	0.112	0.093	0.093	0.188					
	0.114	0.095	0.099	0.152	0.167				
	0.103	0.105	0.102	0.132	0.131	0.144			
	0.065	0.100	0.084	0.073	0.091	0.105	0.233		
	0.121	0.087	0.085	0.110	0.114	0.105	0.067	0.169	
	0.115	0.108	0.105	0.133	0.156	0.128	0.097	0.144	0.353
%	48.8%	50.2%	32.1%	59.9%	39.3%	47.2%	48.6%	61.3%	66.9%
$\hat{\sum} i$									
	0.013	0.001	0.022	0.004	0.036	0.001	0.006	0.004	0.003
%	4.7%	0.4%	6.7%	1.3%	8.3%	0.3%	1.1%	1.5%	0.6%
$\hat{\sum} si$									
	0.001	0.000	0.004	0.000	0.016	0.004	0.002	0.002	0.002
%	0.3%	0.1%	1.1%	0.1%	3.6%	1.4%	0.3%	0.6%	0.3%
$\hat{\sum} pi:s$									
	0.117	0.090	0.151	0.102	0.179	0.116	0.159	0.084	0.085
%	42.3%	33.4%	46.1%	32.6%	42.2%	38.0%	33.2%	30.5%	16.1%
G	0.565	0.847	0.865	0.632	0.652	0.790	0.844	0.625	0.816
Phi	0.494	0.843	0.808	0.609	0.536	0.784	0.828	0.600	0.809
$\sigma^2(\delta)$	0.008	0.008	0.007	0.011	0.014	0.010	0.015	0.010	0.019
$\sigma^2(\Delta)$	0.011	0.008	0.011	0.012	0.023	0.011	0.017	0.011	0.020



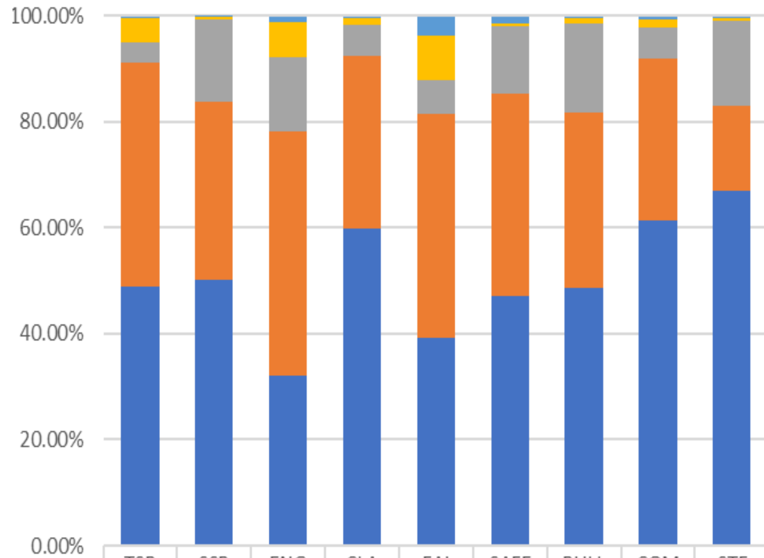
	TSR	SSR	ENG	CLA	FAI	SAFE	BULL	COM	STF
si	0.99%	0.70%	2.71%	0.69%	4.15%	1.17%	0.37%	0.78%	0.22%
i	0.060	0.003	0.052	0.005	0.108	0.001	0.023	0.019	0.004
s	6.34%	17.83%	13.02%	11.07%	8.29%	15.01%	16.07%	10.29%	20.24%
p:s	42.77%	48.24%	32.07%	56.33%	36.81%	48.12%	45.44%	57.57%	63.23%
pi:s	43.87%	32.92%	47.04%	31.40%	39.90%	35.64%	35.81%	29.51%	15.95%

Figure 4.4: Proportions of each variance component to the total variance for elementary school teachers in 2015.



	TSR	SSR	ENG	CLA	FAI	SAFE	BULL	COM	STF
si	0.50%	0.30%	0.70%	0.40%	4.00%	1.10%	0.30%	0.40%	0.20%
i	3.70%	0.60%	4.30%	0.90%	8.80%	0.10%	4.90%	1.50%	0.40%
s	7.60%	19.60%	16.50%	13.10%	12.00%	18.90%	16.50%	7.90%	22.30%
pi:s	39.90%	32.30%	46.00%	30.40%	38.60%	33.80%	35.10%	32.30%	15.40%
p:s	48.30%	47.10%	32.40%	55.20%	36.60%	46.30%	43.20%	57.90%	61.70%

Figure 4.5: Proportions of each variance component to the total variance for elementary school teachers in 2016.



	TSR	SSR	ENG	CLA	FAI	SAFE	BULL	COM	STF
si	0.30%	0.10%	1.10%	0.10%	3.60%	1.40%	0.30%	0.60%	0.30%
i	4.70%	0.40%	6.70%	1.30%	8.30%	0.30%	1.10%	1.50%	0.60%
s	3.70%	15.70%	13.80%	5.80%	6.30%	12.90%	16.60%	5.90%	15.90%
pi:s	42.30%	33.40%	46.10%	32.60%	42.20%	38.00%	33.20%	30.50%	16.10%
p:s	48.80%	50.20%	32.10%	59.90%	39.30%	47.20%	48.60%	61.30%	66.90%

Figure 4.6: Proportions of each variance component to the total variance for elementary school teachers in 2017.

For elementary school parent sample, the results of the G studies were quite stable from 2015 to 2017, as presented from Figure 4.7 to Table 4.9. The parent within school facet explained the largest proportion of the total variance in all the six subscales, ranging from 63.9% in teacher-student relations in 2015 to 75.4% in clarity of behavioral expectations in 2017. Residual variance was the second largest variance component, ranging from 22.7% in clarity of behavioral expectations in 2017 to 29.0% in teacher-student relations in 2015. The school facet explained a certain amount of the total variance, with a minimum of 1.53% in teacher-student relations in 2017 and a maximum of 2.8% in teacher-staff relations in 2016. The item facet accounted for a negligible proportion of the total variance in each subscale, ranging from 0.04% in school safety in 2015 to 1.9% in teacher-student relations in 2016. The interaction between item and school facet explained a minimal amount of the total variance ranging from 0.01% in school safety in 2015 to 0.19% in teacher-home communication in 2017. Tables 4.8 to 4.10 show the estimated covariance components at the parent and school level and correlations for schools across six subscales in elementary schools from 2015 to 2017. Universe score correlations among the school-level subscales were quite high ranging from .81 (student-student relations and teacher home communication in 2017) to .99 (teacher-student relations and clarity of behavioral expectations in 2015). The almost perfect correlations between the six subscales indicate that the subscales might measure redundant aspects of school climate based on parents' perspectives.

Table 4.8: Estimated Variance and covariance component matrices for elementary school parents in 2015

Effect	TSR	SSR	CLA	FAI	SAFE	HCOM
$\hat{\sum} s$						
	0.007	0.941	0.974	0.977	0.949	0.955
	0.009	0.013	0.922	0.918	0.978	0.847
	0.006	0.008	0.006	0.977	0.961	0.881
	0.006	0.008	0.006	0.006	0.947	0.942
	0.008	0.011	0.008	0.007	0.010	0.845
	0.006	0.008	0.006	0.006	0.007	0.006
%	2.1%	3.6%	2.2%	2.0%	3.1%	1.7%
$\hat{\sum} p:s$						
	0.199					
	0.165	0.250				
	0.180	0.156	0.213			
	0.190	0.163	0.198	0.212		
	0.185	0.182	0.193	0.195	0.221	
	0.205	0.161	0.178	0.188	0.180	0.239
%	63.8%	69.9%	74.1%	68.0%	72.0%	68.1%
$\hat{\sum} i$						
	0.006	0.0001	0.0001	0.001	0.0001	0.002
%	1.8%	0.1%	0.1%	0.3%	0.04%	0.5%
$\hat{\sum} si$						
	0.0004	0.0002	0.0001	0.0001	0.0000	0.0004
%	0.13%	0.05%	0.03%	0.03%	0.01%	0.12%
$\hat{\sum} pi:s$						
	0.099	0.094	0.067	0.092	0.076	0.103
%	31.9%	26.3%	23.4%	29.5%	24.7%	29.4%
G	0.738	0.818	0.721	0.716	0.788	0.682
Phi	0.654	0.814	0.713	0.692	0.786	0.647
$\sigma^2(\delta)$	0.002	0.003	0.002	0.003	0.003	0.003
$\sigma^2(\Delta)$	0.004	0.003	0.003	0.003	0.003	0.003

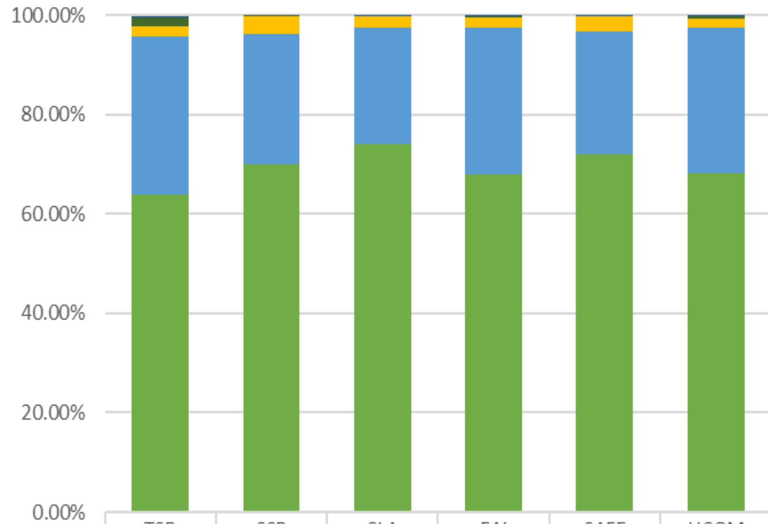
Note. TSR= teacher-student relations, SSR= student-student relations, ENG = school-wide engagement, CLA= clarity of behavioral expectations, FAI = fairness of school rules, SAFE= school safety, HCOM= teacher-home communication.

Table 4.9: Estimated Variance and covariance component matrices for elementary school parents in 2016

Effect	TSR	SSR	CLA	FAI	SAFE	HCOM
$\hat{\sum} s$						
	0.007	0.955	0.996	0.996	0.959	0.971
	0.009	0.012	0.939	0.935	0.965	0.918
	0.007	0.008	0.007	0.992	0.961	0.950
	0.007	0.009	0.007	0.007	0.941	0.980
	0.009	0.011	0.008	0.008	0.011	0.892
	0.006	0.008	0.006	0.006	0.007	0.006
%	2.3%	3.4%	2.3%	2.2%	3.4%	1.6%
$\hat{\sum} p:s$						
	0.210					
	0.163	0.235				
	0.185	0.155	0.210			
	0.196	0.160	0.205	0.220		
	0.187	0.183	0.192	0.196	0.219	
	0.218	0.162	0.186	0.198	0.186	0.248
%	65.1%	70.0%	72.2%	68.1%	69.4%	69.1%
$\hat{\sum} i$						
	0.006	0.002	0.001	0.001	0.001	0.002
%	1.9%	0.5%	0.3%	0.3%	0.3%	0.4%
$\hat{\sum} si$						
	0.0003	0.0002	0.0001	0.0003	0.0002	0.0006
%	0.08%	0.06%	0.03%	0.09%	0.07%	0.17%
$\hat{\sum} pi:s$						
	0.098	0.087	0.073	0.094	0.084	0.103
%	30.5%	25.9%	25.0%	29.1%	26.6%	28.6%
G	0.734	0.795	0.719	0.713	0.787	0.641
Phi	0.655	0.777	0.702	0.694	0.767	0.612
$\sigma^2(\delta)$	0.003	0.003	0.003	0.003	0.003	0.003
$\sigma^2(\Delta)$	0.004	0.003	0.003	0.003	0.003	0.004

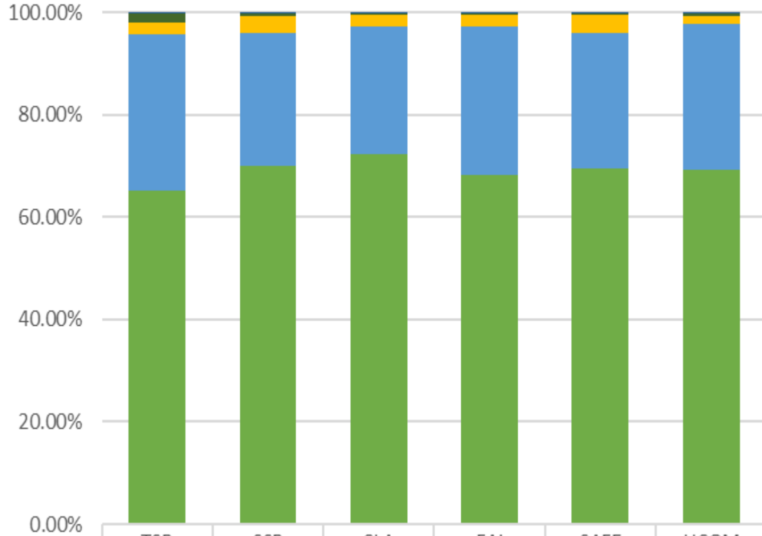
Table 4.10: Estimated Variance and covariance component matrices for elementary school parents in 2017

Effect	TSR	SSR	CLA	FAI	SAFE	HCOM
$\hat{\sum}s$						
	0.006	0.954	0.968	0.975	0.944	0.931
	0.007	0.009	0.931	0.946	0.965	0.812
	0.005	0.006	0.005	0.993	0.977	0.859
	0.006	0.007	0.005	0.006	0.983	0.889
	0.006	0.008	0.006	0.007	0.008	0.833
	0.005	0.006	0.005	0.005	0.006	0.006
%	1.7%	2.5%	1.7%	1.8%	2.7%	1.5%
$\hat{\sum}p:s$						
	0.220					
	0.186	0.261				
	0.196	0.181	0.219			
	0.205	0.187	0.210	0.223		
	0.198	0.202	0.205	0.205	0.223	
	0.227	0.180	0.196	0.204	0.195	0.260
%	67.5%	73.5%	75.4%	69.8%	73.1%	70.6%
$\hat{\sum}i$						
	0.005	0.001	0.000	0.002	0.001	0.002
%	1.6%	0.2%	0.1%	0.7%	0.1%	0.4%
$\hat{\sum}si$						
	0.0001	0.0002	0.0001	0.0004	0.0001	0.0007
%	0.04%	0.06%	0.04%	0.12%	0.03%	0.19%
$\hat{\sum}pi:s$						
	0.094	0.084	0.066	0.088	0.073	0.100
%	29.0%	23.6%	22.7%	27.5%	23.8%	27.1%
G	0.624	0.695	0.595	0.618	0.705	0.573
Phi	0.557	0.688	0.589	0.583	0.695	0.548
$\sigma^2(\delta)$	0.003	0.004	0.003	0.004	0.004	0.004
$\sigma^2(\Delta)$	0.004	0.004	0.003	0.004	0.004	0.005



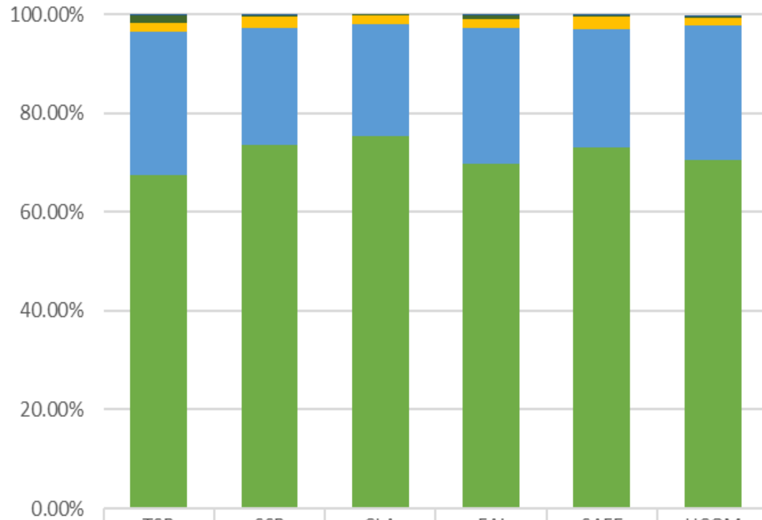
	TSR	SSR	CLA	FAI	SAFE	HCOM
si	0.13%	0.05%	0.03%	0.03%	0.01%	0.12%
i	1.80%	0.10%	0.10%	0.30%	0.04%	0.50%
s	2.10%	3.60%	2.20%	2.00%	3.10%	1.70%
pi:s	31.90%	26.30%	23.40%	29.50%	24.70%	29.40%
p:s	63.80%	69.90%	74.10%	68.00%	72.00%	68.10%

Figure 4.7: Proportions of each variance component to the total variance for elementary school parents in 2015.



	TSR	SSR	CLA	FAI	SAFE	HCOM
si	0.08%	0.06%	0.03%	0.09%	0.07%	0.17%
i	1.90%	0.50%	0.30%	0.30%	0.30%	0.40%
s	2.30%	3.40%	2.30%	2.20%	3.40%	1.60%
pi:s	30.50%	25.90%	25.00%	29.10%	26.60%	28.60%
p:s	65.10%	70.00%	72.20%	68.10%	69.40%	69.10%

Figure 4.8: Proportions of each variance component to the total variance for elementary school parents in 2016.



	TSR	SSR	CLA	FAI	SAFE	HCOM
si	0.04%	0.06%	0.04%	0.12%	0.03%	0.19%
i	1.60%	0.20%	0.10%	0.70%	0.10%	0.40%
s	1.70%	2.50%	1.70%	1.80%	2.70%	1.50%
pi:s	29.00%	23.60%	22.70%	27.50%	23.80%	27.10%
p:s	67.50%	73.50%	75.40%	69.80%	73.10%	70.60%

Figure 4.9: Proportions of each variance component to the total variance for elementary school parents in 2017.

4.1.2 Results of G Studies in Middle Schools

The G studies results of middle school students were consistent from 2015 to 2017, as presented in Figure 4.10 to 4.12. The two greatest variance components were the student within the school facet and the residual variance. The student within the school facet accounted for the largest proportion of the total variance in five subscales, ranging from 37.2% in school-wide engagement in 2015 to 53.9% in teacher-student relations in 2015 and 2017. The residual variance was the greatest variance component in two subscales: clarity of expectations and school-wide engagement, and the second largest variance component in five subscales, ranging from 40.1% in teacher-student relations in 2016 to 57.4% in school-wide engagement in 2015 and 2017. The large proportion of student within school variance indicates that the DSCS was able to discriminate school climate ratings within schools reliably. The third largest variance component was the school facet, which accounted for 1.8% to 7.8% of the total variance in clarity of behavior expectations in 2015 and school safety in 2015. Item facet explains a certain amount of the total variance, ranging from 0.2% in school-wide bullying in 2017 to 4.7% in school safety in 2017. The variance of the school-item interaction term was minimal (less than 1.5% of the total variance). The correlations between the subscales were from moderate to high in middle school student group, with the lowest in .54 (school-wide bullying and fairness of school rules in 2016) and highest .99 (school safety and school-wide engagement in 2016).

Table 4.11: Estimated Variance and covariance component matrices for middle school students in 2015

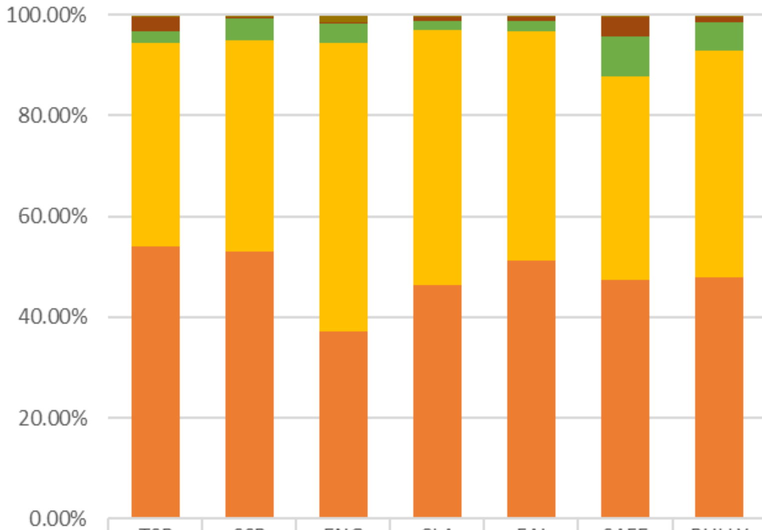
Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULL
$\hat{\sum} s$							
	0.013	0.898	0.945	0.940	0.769	0.899	0.842
	0.016	0.026	0.983	0.863	0.747	0.934	0.979
	0.015	0.023	0.021	0.925	0.820	0.978	0.911
	0.010	0.013	0.013	0.009	0.626	0.949	0.822
	0.010	0.013	0.013	0.007	0.013	0.714	0.641
	0.022	0.032	0.030	0.019	0.017	0.046	0.885
	0.021	0.034	0.029	0.017	0.016	0.041	0.048
%	2.1%	4.4%	3.6%	1.7%	2.1%	7.8%	5.6%
$\hat{\sum} p:s$							
	0.316						
	0.179	0.303					
	0.174	0.210	0.212				
	0.162	0.135	0.131	0.238			
	0.237	0.154	0.159	0.174	0.303		
	0.186	0.217	0.184	0.151	0.171	0.277	
	0.083	0.189	0.108	0.062	0.063	0.159	0.403
%	53.9%	53.0%	37.2%	46.4%	51.1%	47.4%	47.8%
$\hat{\sum} i$							
	0.018	0.002	0.003	0.004	0.005	0.023	0.008
%	3.0%	0.3%	0.5%	0.8%	0.8%	3.8%	0.9%
$\hat{\sum} si$							
	0.001	0.001	0.007	0.001	0.002	0.002	0.003
%	0.1%	0.1%	1.2%	0.2%	0.3%	0.3%	0.4%
$\hat{\sum} pi{:}s$							
	0.238	0.240	0.327	0.259	0.269	0.236	0.379
%	40.6%	41.9%	57.3%	50.6%	45.5%	40.4%	45.1%
G	0.896	0.951	0.897	0.874	0.881	0.959	0.950
Phi	0.714	0.935	0.871	0.794	0.810	0.828	0.914
$\sigma^2(\delta)$	0.001	0.001	0.002	0.001	0.002	0.002	0.003
$\sigma^2(\Delta)$	0.005	0.002	0.003	0.002	0.003	0.009	0.004

Table 4.12: Estimated Variance and covariance component matrices for middle school students in 2016

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULL
$\hat{\sum} s$							
	0.017	0.844	0.944	0.973	0.892	0.891	0.764
	0.018	0.026	0.970	0.740	0.679	0.942	0.966
	0.016	0.020	0.017	0.885	0.782	0.988	0.871
	0.012	0.011	0.011	0.009	0.892	0.859	0.649
	0.015	0.014	0.013	0.010	0.015	0.736	0.541
	0.023	0.030	0.026	0.016	0.018	0.039	0.917
	0.020	0.031	0.023	0.012	0.014	0.037	0.041
%	3.0%	4.5%	3.1%	1.7%	2.6%	6.9%	5.1%
$\hat{\sum} p:s$							
	0.310						
	0.164	0.288					
	0.154	0.202	0.198				
	0.155	0.127	0.122	0.228			
	0.230	0.137	0.137	0.170	0.297		
	0.170	0.205	0.171	0.140	0.152	0.265	
	0.070	0.183	0.100	0.061	0.054	0.151	0.389
%	53.8%	51.4%	36.1%	46.2%	51.4%	46.7%	48.6%
$\hat{\sum} i$							
	0.016	0.002	0.012	0.004	0.003	0.018	0.008
%	2.8%	0.3%	2.2%	0.7%	0.5%	3.1%	0.9%
$\hat{\sum} si$							
	0.001	0.001	0.008	0.001	0.001	0.005	0.004
%	0.1%	0.1%	1.4%	0.2%	0.2%	0.8%	0.5%
$\hat{\sum} pi:s$							
	0.231	0.243	0.312	0.251	0.261	0.239	0.358
%	40.1%	43.4%	56.9%	50.9%	45.1%	42.3%	44.7%
G	0.943	0.960	0.896	0.889	0.927	0.941	0.949
Phi	0.800	0.948	0.809	0.815	0.885	0.824	0.909
$\sigma^2(\delta)$	0.001	0.001	0.002	0.001	0.001	0.002	0.002
$\sigma^2(\Delta)$	0.004	0.001	0.004	0.002	0.002	0.008	0.004

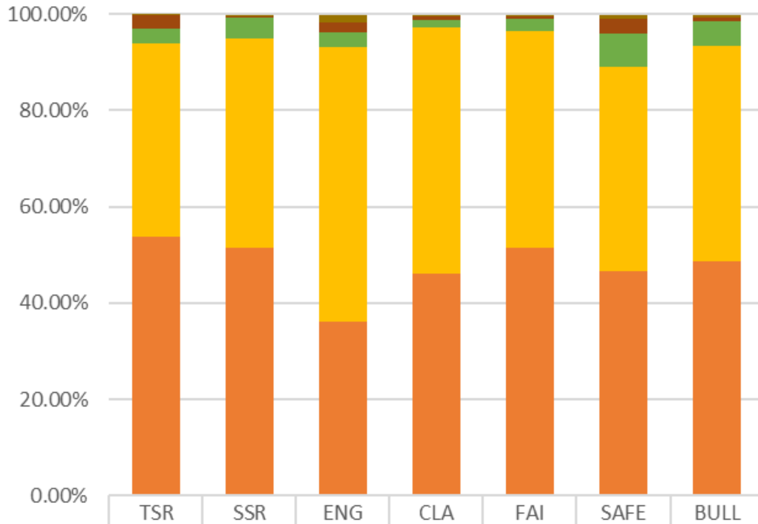
Table 4.13: Estimated Variance and covariance component matrices for middle school students in 2017

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULL
$\hat{\sum} s$							
	0.012	0.676	0.912	0.865	0.941	0.843	0.567
	0.010	0.018	0.909	0.574	0.639	0.857	0.972
	0.011	0.013	0.012	0.747	0.803	0.906	0.790
	0.006	0.005	0.005	0.004	0.887	0.805	0.557
	0.010	0.008	0.008	0.005	0.009	0.820	0.585
	0.015	0.018	0.015	0.008	0.012	0.025	0.821
	0.011	0.023	0.015	0.006	0.009	0.022	0.030
%	2.2%	3.3%	2.2%	0.9%	1.5%	4.8%	4.0%
$\hat{\sum} p:s$							
	0.296						
	0.164	0.276					
	0.152	0.197	0.195				
	0.154	0.125	0.121	0.225			
	0.217	0.138	0.136	0.168	0.281		
	0.163	0.189	0.159	0.135	0.145	0.233	
	0.091	0.196	0.115	0.074	0.074	0.157	0.369
%	53.8%	51.3%	37.4%	48.2%	50.0%	45.6%	49.8%
$\hat{\sum} i$							
	0.015	0.002	0.011	0.003	0.004	0.024	0.001
%	2.6%	0.3%	2.1%	0.6%	0.6%	4.7%	0.1%
$\hat{\sum} si$							
	0.001	0.001	0.004	0.001	0.002	0.003	0.002
%	0.1%	0.1%	0.8%	0.1%	0.2%	0.5%	0.3%
$\hat{\sum} pi:s$							
	0.226	0.241	0.299	0.234	0.267	0.226	0.338
%	41.9%	44.7%	57.3%	49.9%	47.5%	44.3%	45.6%
G	0.932	0.956	0.901	0.843	0.883	0.937	0.941
Phi	0.760	0.937	0.789	0.732	0.811	0.717	0.929
$\sigma^2(\delta)$	0.001	0.001	0.001	0.001	0.001	0.002	0.002
$\sigma^2(\Delta)$	0.004	0.001	0.003	0.002	0.002	0.010	0.002



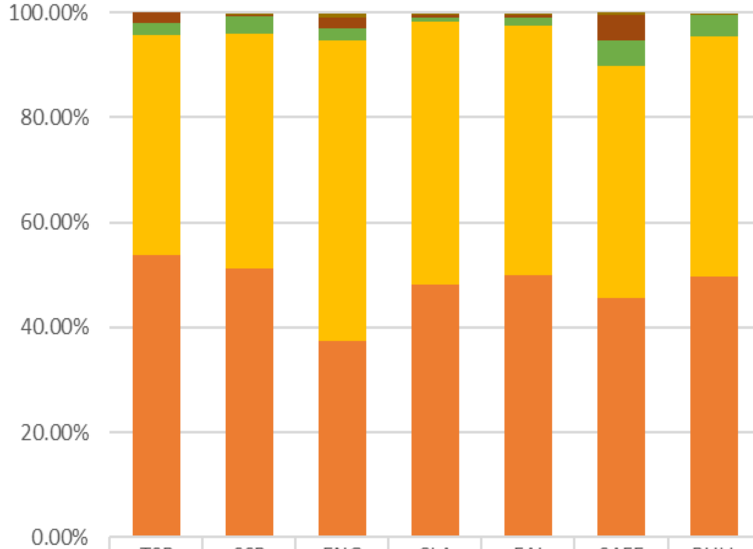
	TSR	SSR	ENG	CLA	FAI	SAFE	BULLY
si	0.10%	0.10%	1.20%	0.20%	0.30%	0.30%	0.40%
i	3.00%	0.30%	0.50%	0.80%	0.80%	3.80%	0.90%
s	2.10%	4.40%	3.60%	1.70%	2.10%	7.80%	5.60%
pi:s	40.60%	41.90%	57.30%	50.60%	45.50%	40.40%	45.10%
p:s	53.90%	53.00%	37.20%	46.40%	51.10%	47.40%	47.80%

Figure 4.10: Proportions of each variance component to the total variance for middle school students in 2015.



	TSR	SSR	ENG	CLA	FAI	SAFE	BULL
si	0.10%	0.10%	1.40%	0.20%	0.20%	0.80%	0.50%
i	2.80%	0.30%	2.20%	0.70%	0.50%	3.10%	0.90%
s	3.00%	4.50%	3.10%	1.70%	2.60%	6.90%	5.10%
pi:s	40.10%	43.40%	56.90%	50.90%	45.10%	42.30%	44.70%
p:s	53.80%	51.40%	36.10%	46.20%	51.40%	46.70%	48.60%

Figure 4.11: Proportions of each variance component to the total variance for middle school students in 2016.



	TSR	SSR	ENG	CLA	FAI	SAFE	BULL
si	0.10%	0.10%	0.80%	0.10%	0.20%	0.50%	0.30%
i	2.60%	0.30%	2.10%	0.60%	0.60%	4.70%	0.10%
s	2.20%	3.30%	2.20%	0.90%	1.50%	4.80%	4.00%
pi:s	41.90%	44.70%	57.30%	49.90%	47.50%	44.30%	45.60%
p:s	53.80%	51.30%	37.40%	48.20%	50.00%	45.60%	49.80%

Figure 4.12: Proportions of each variance component to the total variance for middle school students in 2017.

The G studies results of middle school teachers are presented in Figure 4.13 to 4.15, which were similar to the elementary school teacher results. The two largest variance components are the teacher within the school facet and the residual variance facet. The teacher within school facet accounted for the largest proportion of the total variance in six subscales, and the residual variance facet was the largest variance component in three subscales. School facet was the third largest variance component, ranging from 2.95% in teacher-home communication in 2016 to 33.54% in school safety in 2015. Item facet explains a certain proportion of the total variance, ranging from .05% in school safety 2015 to 13.72% in school-wide engagement in 2017. The interaction effect between school and item facet accounted for a small proportion of the total variance, ranging from .06% in student-student relations in 2016 to 5.58% in fairness of school rules in 2015. The correlations between the subscales were also quite high in middle schools, with the lowest in 0.466 (teacher-student relations and teacher-staff relations in 2017) and highest in 1.00 between several subscales. The high correlations among subscales across years suggest that items in different subscales measure overlapping aspects of middle school teachers' perspectives of school climate.

Table 4.14: Estimated Variance and covariance component matrices for middle school teachers in 2015

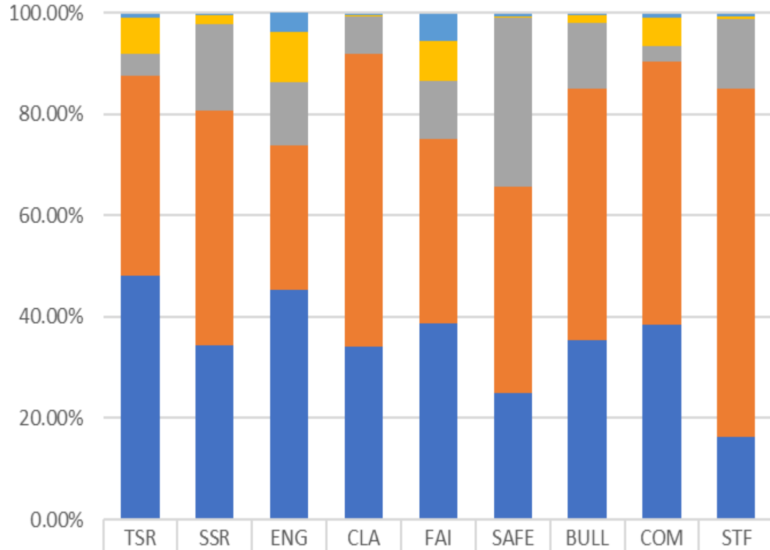
Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULL	COM	STF
$\hat{\sum} s$									
	0.013	0.906	0.983	0.787	0.836	0.713	0.841	0.972	0.745
	0.025	0.060	0.969	0.918	0.905	0.941	1.003	0.854	0.682
	0.027	0.057	0.057	0.828	0.781	0.821	0.901	0.957	0.696
	0.017	0.042	0.037	0.035	0.986	0.948	1.000	0.756	0.740
	0.023	0.055	0.046	0.046	0.061	0.985	0.953	0.740	0.608
	0.033	0.094	0.080	0.072	0.099	0.166	0.972	0.642	0.543
	0.023	0.061	0.053	0.046	0.058	0.098	0.061	0.763	0.707
	0.010	0.020	0.022	0.013	0.017	0.025	0.018	0.009	0.729
	0.025	0.049	0.049	0.041	0.044	0.065	0.052	0.021	0.087
%	4.3%	17.1%	12.5%	7.4%	11.6%	33.5%	12.9%	3.1%	13.8%
$\hat{\sum} p:s$									
	0.116								
	0.057	0.162							
	0.051	0.118	0.130						
	0.093	0.096	0.094	0.271					
	0.100	0.082	0.081	0.181	0.191				
	0.070	0.135	0.113	0.136	0.113	0.201			
	0.041	0.106	0.089	0.069	0.060	0.130	0.233		
	0.106	0.048	0.050	0.096	0.092	0.063	0.035	0.151	
	0.107	0.118	0.118	0.180	0.159	0.145	0.118	0.140	0.433
%	39.4%	46.3%	28.4%	57.6%	36.3%	40.6%	49.6%	52.1%	68.7%
$\hat{\sum} i$									
	0.021	0.006	0.046	0.002	0.041	0.000	0.008	0.016	0.003
%	7.0%	1.7%	9.9%	0.4%	7.7%	0.05%	1.6%	5.6%	0.5%
$\hat{\sum} si$									
	0.003	0.002	0.017	0.001	0.029	0.004	0.002	0.002	0.004
%	0.9%	0.4%	3.7%	0.2%	5.5%	0.7%	0.3%	0.7%	0.5%
$\hat{\sum} pi:s$									
	0.142	0.120	0.208	0.161	0.204	0.124	0.166	0.111	0.103
%	48.2%	34.3%	45.4%	34.2%	38.7%	25.0%	35.4%	38.3%	16.2%
G	0.682	0.893	0.865	0.747	0.790	0.942	0.852	0.559	0.830
Phi	0.558	0.877	0.776	0.739	0.698	0.942	0.830	0.447	0.824
$\sigma^2(\delta)$	0.006	0.007	0.009	0.012	0.016	0.010	0.011	0.007	0.018
$\sigma^2(\Delta)$	0.010	0.008	0.017	0.012	0.026	0.010	0.012	0.011	0.019

Table 4.15: Estimated Variance and covariance component matrices for middle school teacher in 2016

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULL	COM	STF
$\hat{\sum} s$									
	0.009	0.988	0.921	0.894	0.863	0.865	1.010	0.879	0.615
	0.022	0.052	0.996	0.882	0.846	0.943	1.002	0.939	0.719
	0.020	0.051	0.049	0.881	0.786	0.914	0.978	0.856	0.724
	0.011	0.025	0.025	0.016	1.046	0.781	0.938	0.834	0.819
	0.018	0.042	0.038	0.028	0.046	0.820	0.971	0.861	0.704
	0.023	0.060	0.056	0.027	0.049	0.076	1.028	0.726	0.514
	0.020	0.046	0.044	0.024	0.042	0.058	0.041	0.816	0.601
	0.008	0.020	0.017	0.010	0.017	0.018	0.015	0.008	0.558
	0.020	0.056	0.054	0.035	0.051	0.048	0.041	0.017	0.115
%	3.1%	16.0%	11.9%	4.0%	8.9%	20.4%	9.1%	2.9%	19.4%
$\hat{\sum} p:s$									
	0.129								
	0.049	0.146							
	0.046	0.108	0.117						
	0.092	0.071	0.063	0.227					
	0.091	0.075	0.077	0.155	0.179				
	0.058	0.115	0.097	0.104	0.106	0.177			
	0.036	0.100	0.075	0.050	0.056	0.117	0.214		
	0.109	0.037	0.044	0.097	0.092	0.052	0.034	0.155	
	0.095	0.094	0.104	0.126	0.162	0.128	0.088	0.131	0.370
%	44.1%	44.6%	28.2%	57.9%	34.5%	47.2%	47.9%	54.6%	62.5%
$\hat{\sum} i$									
	0.016	0.005	0.046	0.002	0.044	0.001	0.031	0.013	0.005
%	5.3%	1.6%	11.1%	0.4%	8.4%	0.3%	6.8%	4.6%	0.7%
$\hat{\sum} si$									
	0.002	0.000	0.005	0.002	0.024	0.003	0.002	0.001	0.002
%	0.7%	0.06%	1.3%	0.4%	4.5%	0.9%	0.5%	0.2%	0.3%
$\hat{\sum} pi:s$									
	0.137	0.123	0.195	0.145	0.225	0.116	0.159	0.107	0.100
%	46.6%	37.6%	47.3%	37.0%	43.4%	31.0%	35.5%	37.5%	16.8%
G	0.618	0.900	0.892	0.629	0.769	0.901	0.818	0.570	0.893
Phi	0.511	0.884	0.783	0.619	0.650	0.896	0.709	0.465	0.884
$\sigma^2(\delta)$	0.006	0.006	0.006	0.009	0.014	0.008	0.009	0.006	0.014
$\sigma^2(\Delta)$	0.009	0.007	0.014	0.010	0.025	0.009	0.017	0.010	0.015

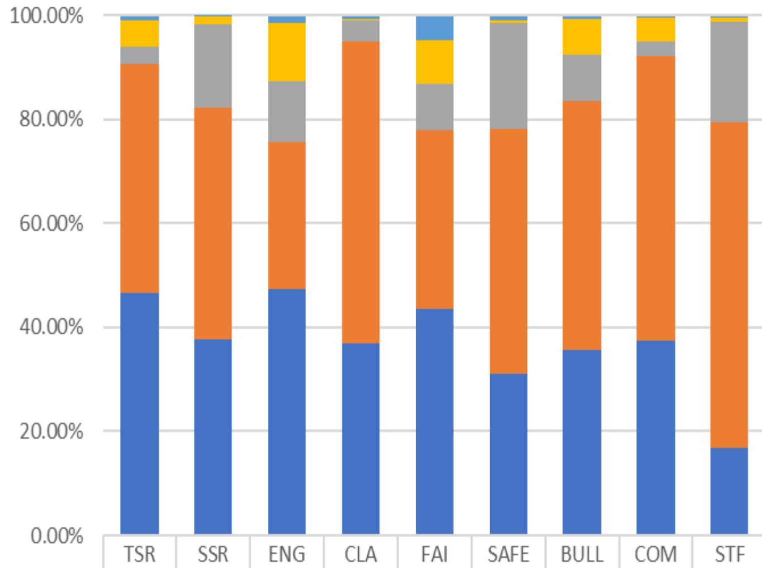
Table 4.16: Estimated Variance and covariance component matrices for middle school teacher in 2017

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULLY	COM	STF
$\hat{\sum}s$									
	0.012	0.729	0.805	0.616	0.748	0.662	0.714	0.965	0.466
	0.019	0.054	1.000	0.703	0.639	0.904	0.939	0.777	0.669
	0.020	0.053	0.052	0.671	0.623	0.875	0.933	0.788	0.605
	0.012	0.029	0.027	0.031	0.883	0.929	0.853	0.812	0.829
	0.014	0.025	0.024	0.026	0.028	0.808	0.809	0.877	0.732
	0.022	0.063	0.059	0.049	0.040	0.089	0.971	0.788	0.757
	0.018	0.051	0.049	0.035	0.032	0.068	0.055	0.856	0.649
	0.011	0.018	0.018	0.014	0.014	0.023	0.020	0.010	0.597
	0.019	0.057	0.050	0.053	0.044	0.082	0.055	0.022	0.132
%	4.1%	17.7%	12.5%	8.3%	6.2%	24.1%	12.8%	3.6%	22.0%
$\hat{\sum}p:s$									
	0.130								
	0.041	0.133							
	0.040	0.098	0.118						
	0.086	0.057	0.069	0.204					
	0.085	0.063	0.065	0.138	0.161				
	0.063	0.104	0.095	0.110	0.109	0.160			
	0.039	0.088	0.066	0.047	0.056	0.094	0.229		
	0.110	0.035	0.037	0.087	0.077	0.059	0.031	0.147	
	0.100	0.100	0.098	0.140	0.150	0.141	0.092	0.116	0.362
%	44.4%	43.5%	28.7%	54.6%	36.1%	43.2%	53.8%	53.8%	60.2%
$\hat{\sum}i$									
	0.014	0.005	0.057	0.002	0.038	0.000	0.002	0.012	0.005
%	4.8%	1.6%	13.7%	0.5%	8.6%	0.07%	0.4%	4.2%	0.9%
$\hat{\sum}si$									
	0.002	0.002	0.005	0.002	0.016	0.006	0.001	0.002	0.001
%	0.6%	0.6%	1.1%	0.4%	3.5%	1.5%	0.1%	0.7%	0.09%
$\hat{\sum}pi:s$									
	0.134	0.111	0.181	0.135	0.203	0.114	0.139	0.102	0.101
%	45.8%	36.4%	43.9%	36.0%	45.5%	30.9%	32.6%	37.5%	16.7%
G	0.724	0.922	0.914	0.819	0.742	0.924	0.877	0.654	0.724
Phi	0.619	0.906	0.783	0.809	0.590	0.924	0.867	0.548	0.619
$\sigma^2(\delta)$	0.005	0.005	0.005	0.007	0.010	0.007	0.008	0.005	0.005
$\sigma^2(\Delta)$	0.008	0.006	0.014	0.007	0.019	0.007	0.008	0.008	0.008



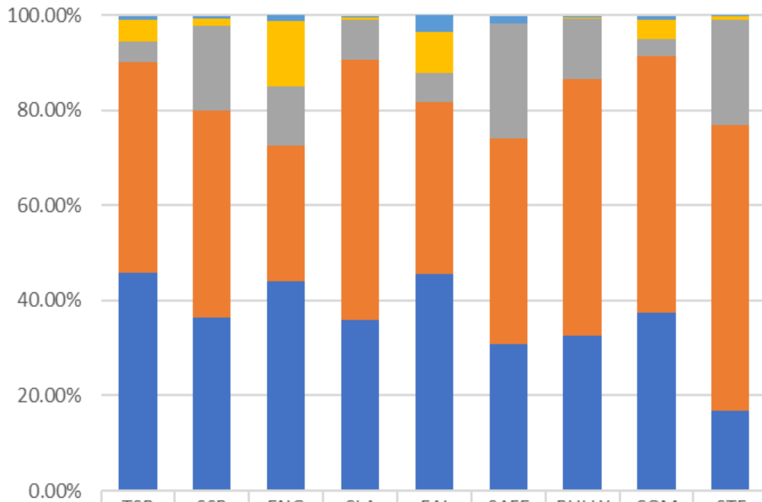
	TSR	SSR	ENG	CLA	FAI	SAFE	BULL	COM	STF
si	0.90%	0.40%	3.70%	0.20%	5.50%	0.70%	0.30%	0.70%	0.50%
i	7.00%	1.70%	9.90%	0.40%	7.70%	0.05%	1.60%	5.60%	0.50%
s	4.30%	17.10%	12.50%	7.40%	11.60%	33.50%	12.90%	3.10%	13.80%
p:s	39.40%	46.30%	28.40%	57.60%	36.30%	40.60%	49.60%	52.10%	68.70%
pi:s	48.20%	34.30%	45.40%	34.20%	38.70%	25.00%	35.40%	38.30%	16.20%

Figure 4.13: Proportions of each variance component to the total variance for middle school teachers in 2015.



	TSR	SSR	ENG	CLA	FAI	SAFE	BULL	COM	STF
si	0.70%	0.06%	1.30%	0.40%	4.50%	0.90%	0.50%	0.20%	0.30%
i	5.30%	1.60%	11.10%	0.40%	8.40%	0.30%	6.80%	4.60%	0.70%
s	3.10%	16.00%	11.90%	4.00%	8.90%	20.40%	9.10%	2.90%	19.40%
p:s	44.10%	44.60%	28.20%	57.90%	34.50%	47.20%	47.90%	54.60%	62.50%
pi:s	46.60%	37.60%	47.30%	37.00%	43.40%	31.00%	35.50%	37.50%	16.80%

Figure 4.14: Proportions of each variance component to the total variance for middle school teachers in 2016.



	TSR	SSR	ENG	CLA	FAI	SAFE	BULLY	COM	STF
si	0.60%	0.60%	1.10%	0.40%	3.50%	1.50%	0.10%	0.70%	0.09%
i	4.80%	1.60%	13.70%	0.50%	8.60%	0.07%	0.40%	4.20%	0.90%
s	4.10%	17.70%	12.50%	8.30%	6.20%	24.10%	12.80%	3.60%	22.00%
p:s	44.40%	43.50%	28.70%	54.60%	36.10%	43.20%	53.80%	53.80%	60.20%
pi:s	45.80%	36.40%	43.90%	36.00%	45.50%	30.90%	32.60%	37.50%	16.70%

Figure 4.15: Proportions of each variance component to the total variance for middle school teachers in 2017.

The G studies results for middle school parents aere presented in Table 4.17 to 4.19, which are similar to the elementary school parent results. The largest variance component was the parent within school facet, ranging from 58.77% in teacher-home communications in 2015 to 71.57% in clarity of behavioral expectations in 2017. The second largest variance component was the residual variance, ranging from 24.29% in student-student relations in 2017 to 36.05% in fairness of school rules in 2016. School facet accounted for a certain proportion of the total variance, ranging from 2.95% in teacher-home communication in 2016 to 8.76% in school safety in 2015. Item facet and the interaction effect between school and item facet accounted for negligible proportions of the total variance (less than two percent). The correlations between the subscales were also quite high in middle schools, with the lowest in 0.33 (school-safety and teacher-home communication in 2016) and highest in 1.00 between five pairs of subscales. The high correlations among subscales across years suggest that items in different subscales measure overlapping aspects of middle school parents' perspectives of school climate.

Table 4.17: Estimated Variance and covariance component matrices for middle school parents in 2015

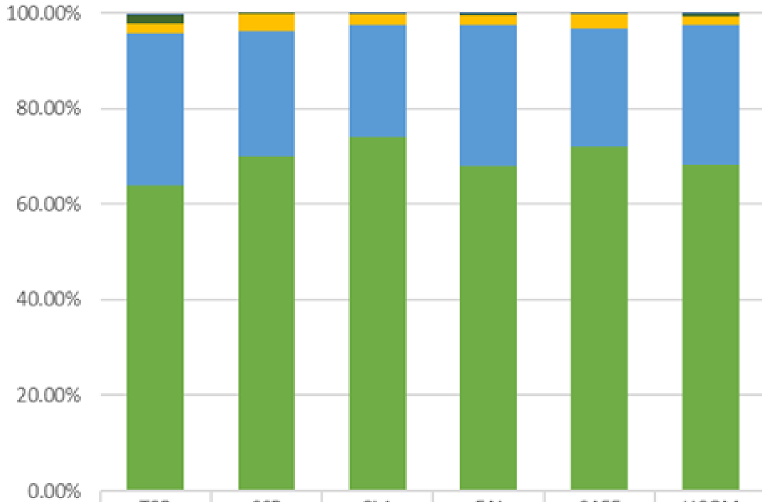
Effect	TSR	SSR	CLA	FAI	SAFE	HCOM
$\hat{\sum} s$						
	0.013	0.988	1.000	0.998	0.949	1.000
	0.018	0.026	0.969	0.929	0.963	0.945
	0.012	0.017	0.012	1.005	0.962	0.966
	0.013	0.017	0.012	0.013	0.938	0.997
	0.020	0.030	0.020	0.020	0.036	0.898
	0.014	0.019	0.013	0.013	0.021	0.015
%	3.5%	5.7%	3.7%	3.6%	8.7%	3.6%
$\hat{\sum} p:s$						
	0.218					
	0.169	0.308				
	0.160	0.129	0.208			
	0.182	0.147	0.176	0.205		
	0.178	0.223	0.167	0.174	0.263	
	0.203	0.145	0.156	0.167	0.161	0.237
%	60.0%	67.1%	65.1%	60.4%	63.2%	58.7%
$\hat{\sum} i$						
	0.006	0.000	0.000	0.001	0.001	0.007
%	1.6%	0.02%	0.05%	0.2%	0.1%	1.6%
$\hat{\sum} si$						
	0.000	0.001	0.000	0.000	0.000	0.001
%	0.06%	0.11%	0.02%	0.01%	0.09%	0.29%
$\hat{\sum} pi:s$						
	0.126	0.124	0.099	0.121	0.116	0.144
%	34.7%	27.0%	31.0%	35.6%	27.8%	35.6%
G	0.811	0.865	0.808	0.815	0.907	0.804
Phi	0.753	0.864	0.806	0.804	0.903	0.736
$\sigma^2(\delta)$	0.003	0.004	0.003	0.003	0.004	0.004
$\sigma^2(\Delta)$	0.004	0.004	0.003	0.003	0.004	0.005

Table 4.18: Estimated Variance and covariance component matrices for middle school parents in 2016

Effect	TSR	SSR	CLA	FAI	SAFE	HCOM
$\hat{\sum} s$						
	0.005	0.874	0.828	0.899	0.616	0.938
	0.008	0.017	1.018	1.034	0.892	0.571
	0.004	0.009	0.005	0.875	0.964	0.541
	0.004	0.009	0.004	0.005	0.857	0.743
	0.007	0.019	0.011	0.009	0.025	0.327
	0.005	0.005	0.003	0.004	0.004	0.005
%	1.3%	3.8%	1.5%	1.3%	5.7%	1.2%
$\hat{\sum} p:s$						
	0.224					
	0.166	0.303				
	0.161	0.139	0.205			
	0.186	0.151	0.179	0.210		
	0.181	0.237	0.163	0.175	0.282	
	0.214	0.153	0.158	0.176	0.169	0.248
%	62.4%	67.8%	66.0%	62.2%	64.5%	61.6%
$\hat{\sum} i$						
	0.006	0.001	0.001	0.001	0.001	0.007
%	1.5%	0.1%	0.08%	0.1%	0.07%	1.8%
$\hat{\sum} si$						
	0.001	0.001	0.001	0.001	0.001	0.001
%	0.1%	0.1%	0.1%	0.1%	0.3%	0.1%
$\hat{\sum} pi:s$						
	0.125	0.125	0.100	0.121	0.128	0.142
%	34.6%	28.0%	32.2%	36.0%	29.2%	35.1%
G	0.591	0.798	0.606	0.584	0.841	0.568
Phi	0.521	0.794	0.601	0.572	0.838	0.469
$\sigma^2(\delta)$	0.003	0.004	0.003	0.003	0.005	0.004
$\sigma^2(\Delta)$	0.004	0.004	0.003	0.003	0.005	0.006

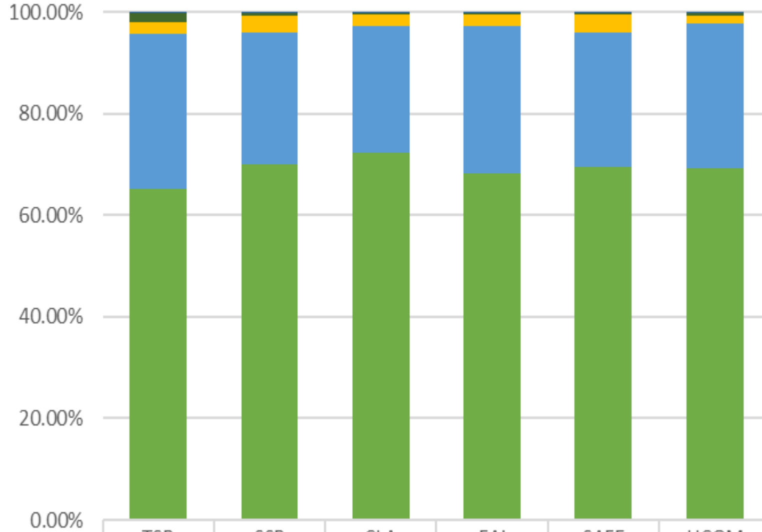
Table 4.19: Estimated Variance and covariance component matrices for middle school parents in 2017

Effect	TSR	SSR	CLA	FAI	SAFE	HCOM
$\hat{\sum} s$						
	0.007	0.946	0.980	1.000	0.827	0.929
	0.011	0.020	0.975	0.965	0.938	0.875
	0.006	0.011	0.006	0.967	0.936	0.864
	0.008	0.012	0.007	0.008	0.900	0.921
	0.009	0.018	0.010	0.011	0.019	0.656
	0.007	0.011	0.006	0.007	0.008	0.007
%	1.8%	4.5%	1.9%	2.3%	5.0%	1.7%
$\hat{\sum} p:s$						
	0.247					
	0.189	0.312				
	0.184	0.156	0.219			
	0.204	0.167	0.200	0.222		
	0.200	0.229	0.183	0.193	0.261	
	0.248	0.184	0.187	0.200	0.194	0.283
%	66.1%	71.0%	71.5%	63.9%	69.6%	66.4%
$\hat{\sum} i$						
	0.006	0.001	0.001	0.002	0.001	0.007
%	1.5%	0.1%	0.02%	0.6%	0.02%	1.5%
$\hat{\sum} si$						
	0	0.001	0.001	0.001	0.001	0.001
%	0%	0.02%	0.02%	0.15%	0.21%	0.08%
$\hat{\sum} pi:s$						
	0.114	0.107	0.081	0.114	0.094	0.128
%	30.5%	24.2%	26.4%	32.8%	25.0%	30.1%
G	0.668	0.827	0.662	0.717	0.828	0.652
Phi	0.601	0.825	0.661	0.685	0.827	0.570
$\sigma^2(\delta)$	0.003	0.004	0.003	0.003	0.004	0.004
$\sigma^2(\Delta)$	0.004	0.004	0.003	0.004	0.004	0.006



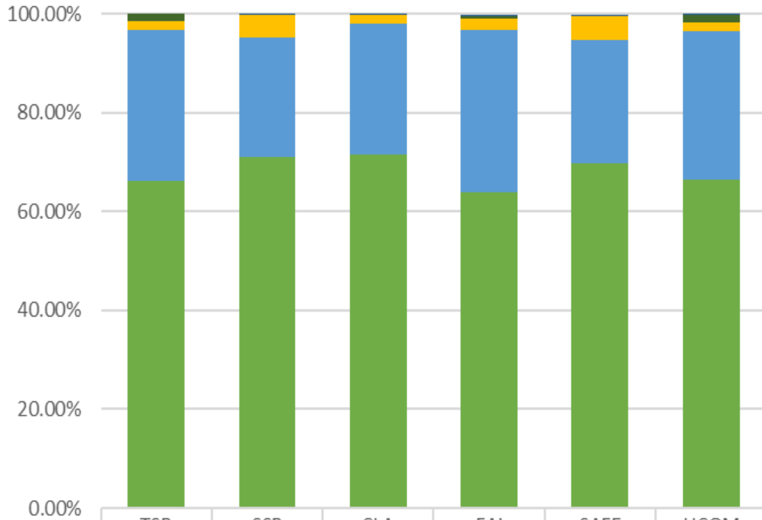
	TSR	SSR	CLA	FAI	SAFE	HCOM
si	0.13%	0.05%	0.03%	0.03%	0.01%	0.12%
i	1.80%	0.10%	0.10%	0.30%	0.04%	0.50%
s	2.10%	3.60%	2.20%	2.00%	3.10%	1.70%
pi:s	31.90%	26.30%	23.40%	29.50%	24.70%	29.40%
p:s	63.80%	69.90%	74.10%	68.00%	72.00%	68.10%

Figure 4.16: Proportions of each variance component to the total variance for middle school parents in 2015.



	TSR	SSR	CLA	FAI	SAFE	HCOM
si	0.08%	0.06%	0.03%	0.09%	0.07%	0.17%
i	1.90%	0.50%	0.30%	0.30%	0.30%	0.40%
s	2.30%	3.40%	2.30%	2.20%	3.40%	1.60%
pi:s	30.50%	25.90%	25.00%	29.10%	26.60%	28.60%
p:s	65.10%	70.00%	72.20%	68.10%	69.40%	69.10%

Figure 4.17: Proportions of each variance component to the total variance for middle school parents in 2016.



	TSR	SSR	CLA	FAI	SAFE	HCOM
si	0%	0.02%	0.02%	0.15%	0.21%	0.08%
i	1.50%	0.10%	0.02%	0.60%	0.02%	1.50%
s	1.80%	4.50%	1.90%	2.30%	5.00%	1.70%
pi:s	30.50%	24.20%	26.40%	32.80%	25.00%	30.10%
p:s	66.10%	71.00%	71.50%	63.90%	69.60%	66.40%

Figure 4.18: Proportions of each variance component to the total variance for middle school parents in 2017.

4.1.3 Results of G Studies in High Schools

Figures 4.19 to 4.21 show the G studies results of high school students, which indicates similar patterns as middle and elementary school students. The student within the school facet accounted for the largest amount of variance in almost all the subscales, except for school-wide engagement (around 35% of the total variance). The second largest variance component was the residual variance including the interaction between students, schools, items within each subscale, and unexplained sources of errors. The percentages range of the residual variance of the total variance were from 37.50% in school safety 2017 to 57.77% in school-wide engagement in 2016. The school variance explained a certain proportion of the total variance, ranging from 1%

in teacher-student relations in 2016 to 9.72% in school safety in 2017. The item facet also accounted for a small amount of variance, ranging from 0.08% in school safety 2017 to 3.35% in teacher-student relations 2017. The variance of the interaction effect between school and item was almost negligible (less than 1.96% of the total variance). The correlations between the subscales varied across subscales, with the lowest as zero (school-wide bullying and fairness of school rules in 2016) and highest as 1 (student-student relations and school-wide engagement in 2016).

Table 4.20: Estimated Variance and covariance component matrices for high school students in 2015

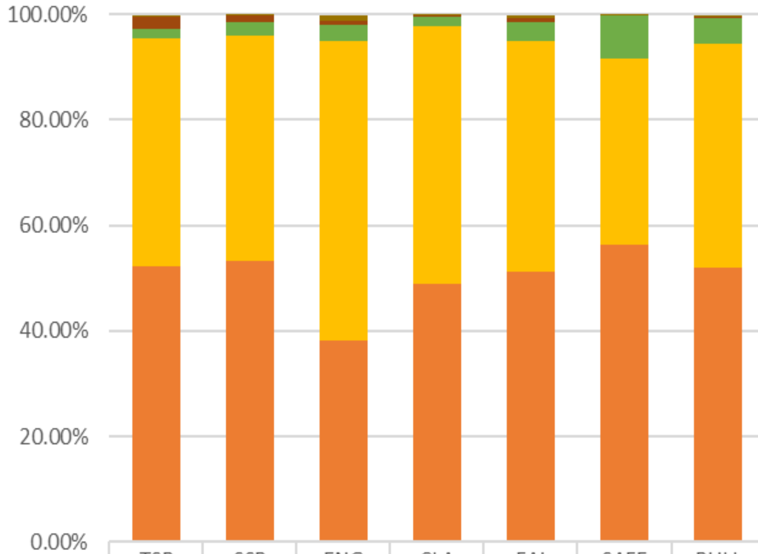
Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULL
$\hat{\sum} s$							
	0.010	0.840	0.801	0.933	0.696	0.761	0.869
	0.010	0.013	0.991	0.847	0.444	0.903	0.913
	0.010	0.014	0.015	0.870	0.501	0.883	0.810
	0.008	0.009	0.010	0.008	0.669	0.864	0.837
	0.009	0.007	0.008	0.008	0.017	0.238	0.335
	0.015	0.020	0.022	0.015	0.006	0.039	0.925
	0.015	0.019	0.018	0.014	0.008	0.032	0.032
%	1.9%	2.7%	2.9%	2.0%	3.5%	8.1%	5.0%
$\hat{\sum} p:s$							
	0.259						
	0.165	0.260					
	0.153	0.196	0.203				
	0.133	0.108	0.106	0.200			
	0.190	0.123	0.124	0.148	0.255		
	0.164	0.198	0.166	0.128	0.137	0.266	
	0.067	0.147	0.081	0.034	0.036	0.145	0.330
%	52.3%	53.2%	38.2%	49.0%	51.3%	56.4%	52.0%
$\hat{\sum} i$							
	0.012	0.005	0.005	0.001	0.005	0.001	0.002
%	2.3%	1.1%	0.9%	0.2%	0.9%	0.1%	0.3%
$\hat{\sum} si$							
	0.001	0.001	0.006	0.000	0.003	0.000	0.001
%	0.2%	0.1%	1.0%	0.1%	0.5%	0.02%	0.1%
$\hat{\sum} pi:s$							
	0.213	0.209	0.301	0.198	0.217	0.165	0.269
%	43.0%	42.7%	56.8%	48.6%	43.6%	35.2%	42.3%
G	0.805	0.855	0.840	0.816	0.862	0.944	0.913
Phi	0.675	0.800	0.798	0.797	0.815	0.940	0.897
$\sigma^2(\delta)$	0.002	0.002	0.003	0.002	0.003	0.002	0.003
$\sigma^2(\Delta)$	0.005	0.003	0.004	0.002	0.004	0.002	0.004

Table 4.21: Estimated Variance and covariance component matrices for high school students in 2016

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULL
$\hat{\sum} s$							
	0.005	0.744	0.731	0.799	0.572	0.733	0.494
	0.006	0.012	1.000	0.779	0.488	0.829	0.671
	0.006	0.012	0.013	0.794	0.499	0.828	0.573
	0.004	0.006	0.007	0.006	0.578	0.719	0.427
	0.004	0.006	0.006	0.005	0.012	0.132	0
	0.008	0.014	0.015	0.008	0.002	0.024	0.738
	0.004	0.009	0.008	0.004	-0.001	0.014	0.015
%	1.0%	2.4%	2.5%	1.4%	2.5%	5.5%	2.5%
$\hat{\sum} p:s$							
	0.254						
	0.163	0.257					
	0.143	0.185	0.186				
	0.128	0.104	0.104	0.196			
	0.187	0.128	0.121	0.148	0.255		
	0.159	0.186	0.154	0.120	0.140	0.242	
	0.067	0.141	0.075	0.043	0.045	0.130	0.312
%	52.5%	53.7%	36.8%	49.5%	50.9%	55.6%	51.9%
$\hat{\sum} i$							
	0.012	0.003	0.009	0.001	0.004	0.001	0.015
%	2.4%	0.7%	1.7%	0.2%	0.8%	0.1%	2.4%
$\hat{\sum} si$							
	0.001	0.001	0.006	0.001	0.003	0.001	0.002
%	0.26%	0.09%	1.11%	0.18%	0.66%	0.20%	0.31%
$\hat{\sum} pi:s$							
	0.211	0.206	0.292	0.193	0.225	0.167	0.257
%	43.6%	43.0%	57.7%	48.6%	45.0%	38.5%	42.7%
G	0.608	0.797	0.801	0.689	0.766	0.884	0.791
Phi	0.467	0.761	0.734	0.670	0.720	0.879	0.665
$\sigma^2(\delta)$	0.003	0.003	0.003	0.003	0.004	0.003	0.004
$\sigma^2(\Delta)$	0.006	0.004	0.005	0.003	0.005	0.003	0.008

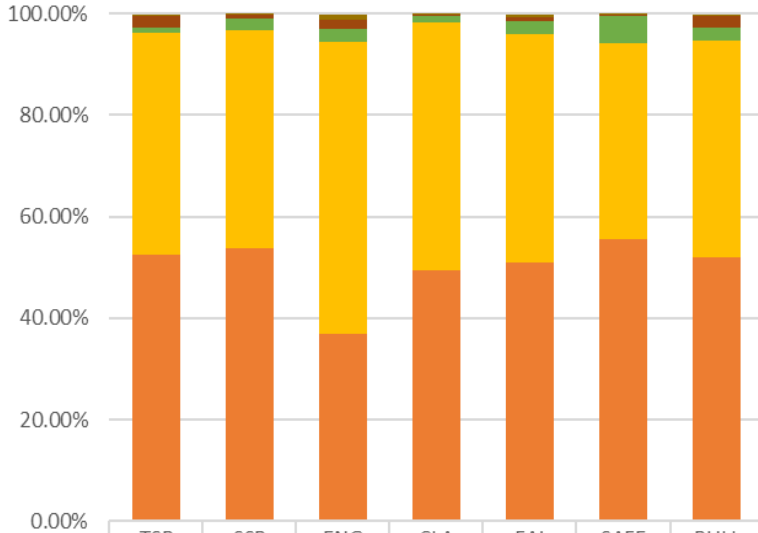
Table 4.22: Estimated Variance and covariance component matrices for high school students in 2017

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULL
$\hat{\sum} s$							
	0.020	0.881	0.890	0.847	0.859	0.836	0.811
	0.021	0.029	0.995	0.871	0.768	0.979	0.928
	0.022	0.029	0.030	0.882	0.820	0.950	0.857
	0.013	0.017	0.017	0.013	0.867	0.890	0.713
	0.018	0.020	0.022	0.015	0.023	0.722	0.650
	0.024	0.034	0.033	0.020	0.022	0.041	0.954
	0.021	0.030	0.028	0.015	0.019	0.036	0.035
%	3.9%	6.0%	6.0%	3.1%	4.4%	9.7%	6.0%
$\hat{\sum} p:s$							
	0.259						
	0.165	0.247					
	0.138	0.175	0.182				
	0.136	0.107	0.103	0.195			
	0.194	0.139	0.126	0.152	0.264		
	0.154	0.173	0.135	0.123	0.136	0.219	
	0.085	0.154	0.089	0.051	0.058	0.132	0.292
%	52.1%	52.0%	36.0%	48.8%	51.5%	52.2%	49.9%
$\hat{\sum} i$							
	0.017	0.002	0.010	0.001	0.003	0.001	0.002
%	3.3%	0.4%	1.8%	0.1%	0.6%	0.08%	0.3%
$\hat{\sum} si$							
	0.001	0.001	0.010	0.001	0.002	0.002	0.003
%	0.16%	0.12%	1.96%	0.22%	0.32%	0.41%	0.45%
$\hat{\sum} pi:s$							
	0.201	0.196	0.273	0.190	0.220	0.157	0.252
%	40.3%	41.3%	54.1%	47.5%	43.0%	37.5%	43.1%
G	0.917	0.946	0.913	0.893	0.915	0.952	0.924
Phi	0.794	0.932	0.872	0.884	0.885	0.950	0.907
$\sigma^2(\delta)$	0.002	0.002	0.003	0.002	0.002	0.002	0.003
$\sigma^2(\Delta)$	0.005	0.002	0.004	0.002	0.003	0.002	0.004



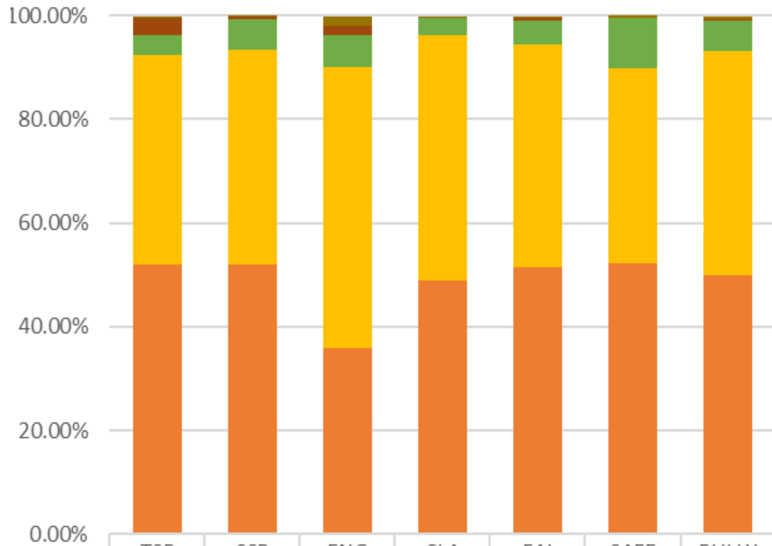
	TSR	SSR	ENG	CLA	FAI	SAFE	BULL
si	0.20%	0.10%	1.00%	0.10%	0.50%	0.02%	0.10%
i	2.30%	1.10%	0.90%	0.20%	0.90%	0.10%	0.30%
s	1.90%	2.70%	2.90%	2.00%	3.50%	8.10%	5.00%
pi:s	43.00%	42.70%	56.80%	48.60%	43.60%	35.20%	42.30%
p:s	52.30%	53.20%	38.20%	49.00%	51.30%	56.40%	52.00%

Figure 4.19: Proportions of each variance component to the total variance for high school students in 2015.



	TSR	SSR	ENG	CLA	FAI	SAFE	BULL
si	0.26%	0.09%	1.11%	0.18%	0.66%	0.20%	0.31%
i	2.40%	0.70%	1.70%	0.20%	0.80%	0.10%	2.40%
s	1.00%	2.40%	2.50%	1.40%	2.50%	5.50%	2.50%
pi:s	43.60%	43.00%	57.70%	48.60%	45.00%	38.50%	42.70%
p:s	52.50%	53.70%	36.80%	49.50%	50.90%	55.60%	51.90%

Figure 4.20: Proportions of each variance component to the total variance for high school students in 2016.



	TSR	SSR	ENG	CLA	FAI	SAFE	BULLY
si	0.16%	0.12%	1.96%	0.22%	0.32%	0.41%	0.45%
i	3.30%	0.40%	1.80%	0.10%	0.60%	0.08%	0.30%
s	3.90%	6.00%	6.00%	3.10%	4.40%	9.70%	6.00%
pi:s	40.30%	41.30%	54.10%	47.50%	43.00%	37.50%	43.10%
p:s	52.10%	52.00%	36.00%	48.80%	51.50%	52.20%	49.90%

Figure 4.21: Proportions of each variance component to the total variance for high school students in 2017.

Figures 4.22 to 4.24 show the G studies results of high school teachers, which indicates similar patterns as middle school teachers. The teachers within the school and residual variance were the two largest variance components from 2015 to 2017. The teacher within school variance ranges from 26.53% in schoolwide engagement in 2016 to 69.29% teacher-staff relations in 2015. The residual variance ranges from 16.31% in teacher-staff relations to 49.18% teacher-student relations in 2016. School facet was the third largest variance component, ranging from 1.34% in teacher-student relations in 2017 to 24.09% in school safety in 2016. In the sample of high school teachers, item facet explains a certain proportion of the total variance, ranging from .16% in school safety in 2015 to 16.92% in school-wide engagement in 2016. The interaction effect between school and item facet also accounted for a small proportion of the total variance, ranging from zero in clarity of expectation in 2016 to 3.13% in fairness of school rules in 2015. The correlations between the subscales were also quite high in the sample of middle school teachers, with the lowest in 0.31(school-wide bullying and teacher-staff relations in 2017) and highest in 1.00 between several subscales. The high correlations among subscales across years are attributable to the fact that items in different subscales measure overlapping aspects of high school teachers' perspectives of school climate.

Table 4.23: Estimated Variance and covariance component matrices for high school teachers in 2015

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULLY	COM	STF
$\hat{\sum} s$									
	0.014	0.870	0.743	0.566	0.493	0.618	0.698	0.962	0.434
	0.018	0.030	0.973	0.732	0.854	0.898	0.949	0.779	0.507
	0.024	0.046	0.077	0.824	0.824	0.823	0.833	0.794	0.604
	0.014	0.026	0.047	0.043	0.859	0.762	0.594	0.639	0.746
	0.010	0.025	0.039	0.030	0.029	1.000	0.936	0.378	0.392
	0.023	0.049	0.073	0.051	0.054	0.102	0.935	0.527	0.338
	0.016	0.032	0.045	0.024	0.031	0.058	0.038	0.564	0.309
	0.016	0.019	0.031	0.019	0.009	0.024	0.015	0.020	0.507
	0.015	0.026	0.050	0.047	0.020	0.032	0.018	0.021	0.090
%	4.6%	9.7%	14.2%	8.8%	6.0%	24.0%	8.8%	6.3%	13.7%
$\hat{\sum} p:s$									
	0.118								
	0.059	0.148							
	0.054	0.114	0.149						
	0.082	0.082	0.099	0.272					
	0.091	0.058	0.068	0.177	0.189				
	0.077	0.120	0.110	0.145	0.117	0.219			
	0.053	0.106	0.095	0.081	0.073	0.151	0.235		
	0.101	0.041	0.049	0.089	0.092	0.066	0.046	0.152	
	0.107	0.115	0.129	0.203	0.176	0.163	0.114	0.136	0.453
%	38.8%	48.6%	27.5%	55.6%	40.1%	51.9%	54.3%	49.0%	69.2%
$\hat{\sum} i$									
	0.022	0.008	0.079	0.001	0.037	0.001	0.009	0.022	0.002
%	7.2%	2.6%	14.5%	0.2%	7.9%	0.1%	1.9%	7.0%	0.3%
$\hat{\sum} si$									
	0.002	0.001	0.009	0.004	0.015	0.001	0.001	0.003	0.002
%	0.6%	0.1%	1.6%	0.7%	3.1%	0.2%	0.2%	1.1%	0.3%
$\hat{\sum} pi:s$									
	0.148	0.118	0.227	0.168	0.202	0.100	0.149	0.113	0.107
%	48.7%	38.7%	41.9%	34.4%	42.7%	23.5%	34.6%	36.5%	16.3%
G	0.712	0.825	0.903	0.780	0.700	0.915	0.792	0.728	0.712
Phi	0.583	0.790	0.783	0.776	0.570	0.913	0.758	0.606	0.583
$\sigma^2(\delta)$	0.006	0.006	0.008	0.012	0.012	0.009	0.010	0.007	0.006
$\sigma^2(\Delta)$	0.010	0.008	0.021	0.013	0.022	0.010	0.012	0.013	0.010

Table 4.24: Estimated Variance and covariance component matrices for high school teachers in 2016

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULLY	COM	STF
$\hat{\sum} s$									
	0.004	0.000	0.000	0.619	0.840	0.145	0.047	1.000	0.593
	0.000	0.012	1.048	0.738	0.660	0.915	0.816	0.169	0.170
	0.000	0.019	0.028	0.598	0.543	0.818	0.859	0.149	0.112
	0.005	0.012	0.014	0.020	0.550	0.661	0.446	0.764	0.513
	0.006	0.009	0.012	0.010	0.017	0.832	0.844	0.405	0.375
	0.002	0.019	0.025	0.017	0.020	0.035	0.887	0.172	0.224
	0.000	0.012	0.019	0.008	0.014	0.021	0.017	0.000	0.151
	0.004	0.001	0.002	0.007	0.003	0.002	0.000	0.004	0.569
	0.010	0.005	0.005	0.020	0.013	0.012	0.005	0.010	0.078
%	1.3%	4.7%	6.1%	4.5%	3.6%	11.7%	4.1%	1.5%	13.9%
$\hat{\sum} p:s$									
	0.117								
	0.051	0.130							
	0.047	0.095	0.119						
	0.073	0.070	0.083	0.257					
	0.081	0.071	0.076	0.161	0.181				
	0.057	0.109	0.093	0.117	0.101	0.167			
	0.034	0.100	0.072	0.057	0.055	0.109	0.210		
	0.084	0.041	0.043	0.078	0.074	0.053	0.026	0.111	
	0.082	0.092	0.103	0.153	0.141	0.118	0.073	0.101	0.367
%	43.9%	49.5%	26.5%	59.6%	40.1%	56.8%	51.4%	43.0%	65.6%
$\hat{\sum} i$									
	0.011	0.005	0.076	0.001	0.029	0.001	0.024	0.022	0.005
%	4.2%	1.7%	16.9%	0.2%	6.3%	0.4%	5.7%	8.3%	0.9%
$\hat{\sum} si$									
	0.003	0.001	0.009	0.001	0.013	0.001	0.001	0.002	0.002
%	1.23%	0.18%	2.10%	0.01%	2.92%	0.42%	0.05%	0.95%	0.44%
$\hat{\sum} pi:s$									
	0.131	0.114	0.217	0.154	0.212	0.090	0.158	0.119	0.106
%	49.1%	43.7%	48.3%	35.5%	46.9%	30.5%	38.6%	46.0%	18.9%
G	0.401	0.710	0.806	0.672	0.602	0.835	0.672	0.433	0.401
Phi	0.319	0.674	0.588	0.666	0.478	0.826	0.544	0.273	0.319
$\sigma^2(\delta)$	0.005	0.005	0.007	0.010	0.011	0.007	0.008	0.005	0.005
$\sigma^2(\Delta)$	0.008	0.006	0.019	0.010	0.018	0.007	0.014	0.011	0.008

Table 4.25: Estimated Variance and covariance component matrices for high school teachers in 2017

Effect	TSR	SSR	ENG	CLA	FAI	SAFE	BULL	COM	STF
$\hat{\sum} s$									
	0.011	0.604	0.535	0.603	0.651	0.548	0.620	0.900	0.441
	0.010	0.027	1.000	0.728	0.790	0.955	0.898	0.517	0.365
	0.013	0.040	0.059	0.712	0.753	0.958	0.901	0.543	0.378
	0.011	0.021	0.031	0.031	1.000	0.693	0.680	0.527	0.914
	0.010	0.019	0.026	0.025	0.020	0.753	0.843	0.458	0.798
	0.013	0.036	0.054	0.029	0.025	0.054	0.996	0.551	0.326
	0.012	0.028	0.041	0.023	0.023	0.043	0.035	0.576	0.382
	0.009	0.008	0.013	0.009	0.006	0.013	0.011	0.010	0.438
	0.013	0.017	0.027	0.047	0.033	0.022	0.021	0.013	0.084
%	3.8%	10.5%	12.8%	8.0%	5.1%	17.8%	8.5%	3.5%	15.0%
$\hat{\sum} p:s$									
	0.125								
	0.060	0.124							
	0.057	0.101	0.130						
	0.093	0.086	0.090	0.231					
	0.100	0.073	0.088	0.152	0.174				
	0.073	0.095	0.091	0.124	0.106	0.151			
	0.056	0.091	0.078	0.082	0.069	0.114	0.244		
	0.099	0.051	0.053	0.088	0.093	0.062	0.049	0.139	
	0.104	0.109	0.129	0.166	0.174	0.129	0.110	0.131	0.376
%	45.3%	48.4%	28.3%	58.7%	44.5%	50.1%	59.3%	50.5%	66.8%
$\hat{\sum} i$									
	0.013	0.005	0.059	0.001	0.020	0.002	0.002	0.019	0.003
%	4.6%	1.8%	13.0%	0.3%	5.0%	0.5%	0.3%	7.0%	0.4%
$\hat{\sum} si$									
	0.002	0.003	0.010	0.001	0.005	0.001	0.002	0.001	0.002
%	0.6%	1.06%	2.1%	0.3%	1.3%	0.3%	0.5%	0.3%	0.3%
$\hat{\sum} pi:s$									
	0.126	0.097	0.199	0.128	0.172	0.094	0.128	0.106	0.097
%	45.4%	38.0%	43.5%	32.6%	43.9%	31.0%	31.1%	38.5%	17.2%
G	0.676	0.842	0.897	0.784	0.714	0.898	0.781	0.644	0.865
Phi	0.581	0.818	0.779	0.779	0.608	0.890	0.772	0.488	0.860
$\sigma^2(\delta)$	0.005	0.005	0.007	0.009	0.008	0.006	0.010	0.005	0.013
$\sigma^2(\Delta)$	0.008	0.006	0.017	0.009	0.013	0.007	0.010	0.010	0.014

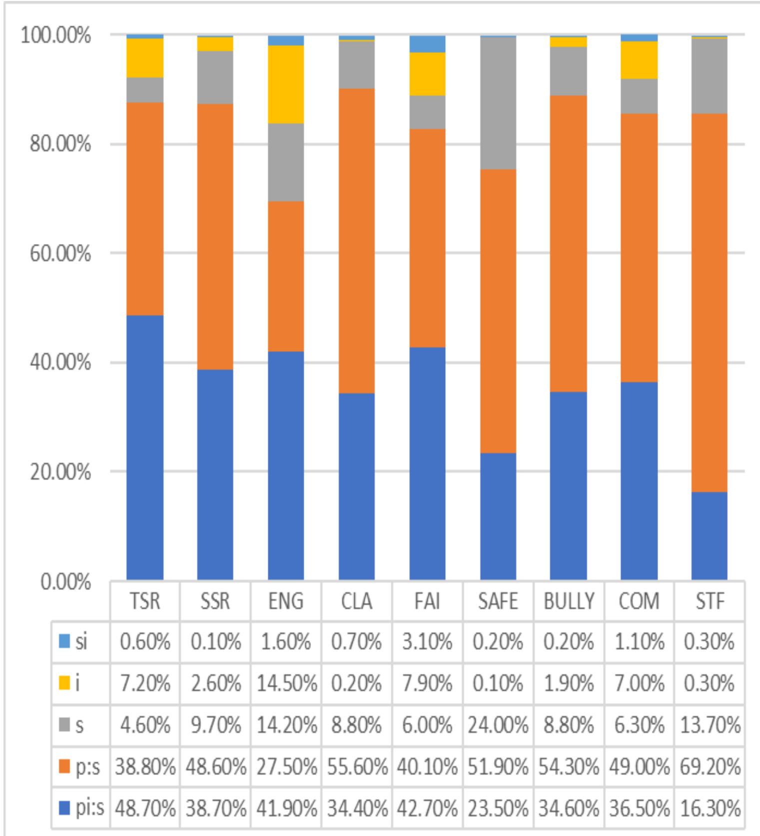
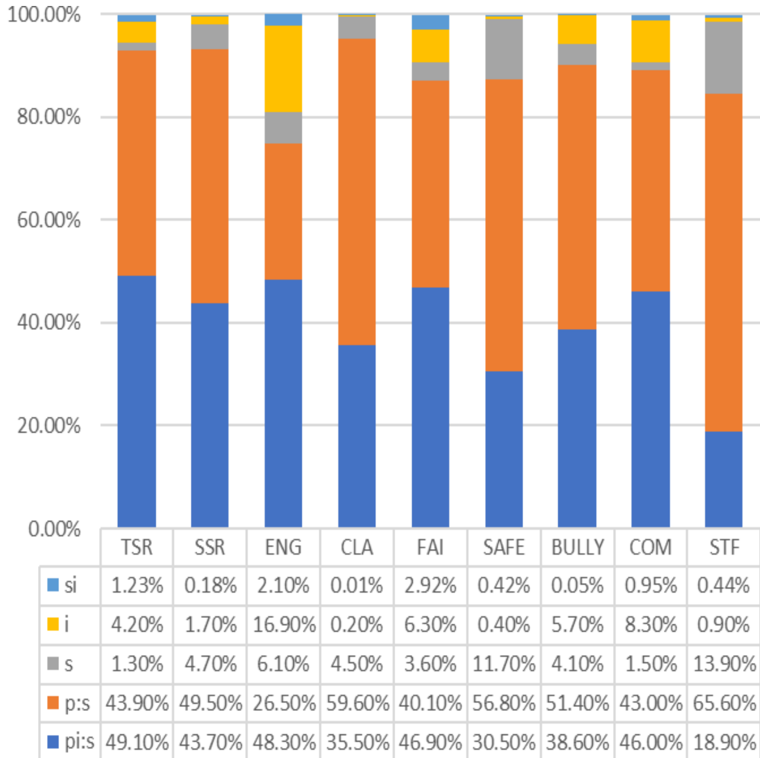
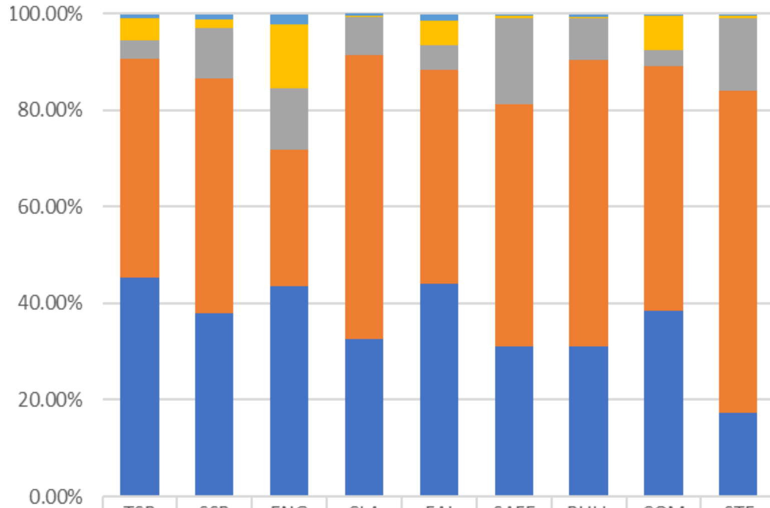


Figure 4.22: Proportions of each variance component to the total variance for high school teachers in 2015



	TSR	SSR	ENG	CLA	FAI	SAFE	BULLY	COM	STF
si	1.23%	0.18%	2.10%	0.01%	2.92%	0.42%	0.05%	0.95%	0.44%
i	4.20%	1.70%	16.90%	0.20%	6.30%	0.40%	5.70%	8.30%	0.90%
s	1.30%	4.70%	6.10%	4.50%	3.60%	11.70%	4.10%	1.50%	13.90%
p:s	43.90%	49.50%	26.50%	59.60%	40.10%	56.80%	51.40%	43.00%	65.60%
pi:s	49.10%	43.70%	48.30%	35.50%	46.90%	30.50%	38.60%	46.00%	18.90%

Figure 4.23: Proportions of each variance component to the total variance for high school teachers in 2016.



	TSR	SSR	ENG	CLA	FAI	SAFE	BULL	COM	STF
si	0.60%	1.06%	2.10%	0.30%	1.30%	0.30%	0.50%	0.30%	0.30%
i	4.60%	1.80%	13.00%	0.30%	5.00%	0.50%	0.30%	7.00%	0.40%
s	3.80%	10.50%	12.80%	8.00%	5.10%	17.80%	8.50%	3.50%	15.00%
p:s	45.30%	48.40%	28.30%	58.70%	44.50%	50.10%	59.30%	50.50%	66.80%
pi:s	45.40%	38.00%	43.50%	32.60%	43.90%	31.00%	31.10%	38.50%	17.20%

Figure 4.24: Proportions of each variance component to the total variance for high school teachers in 2017.

Figures 4.25 to 4.27 show the G studies results of high school parents, which indicates similar patterns as middle school parents. The parents within the school and residual variance were the two largest variance components from 2015 to 2017. The parent within school variance ranges from 21.35% in fairness of school rules in 2015 to 70.65% student-student relations in 2017. The residual variance ranges from 20.2% in clarity of behavioral expectations in 2016 to 61.06% in fairness of school rules in 2015. School facet in high school parents sample explained more proportion of the total variance than it's in elementary and middle school parents sample in 2015 and 2017, ranging from 6.46% in teacher-home communication in 2017 to 17.84% in student-student relations in 2015. In the sample of high school parents, item facet explained a certain proportion of the total variance, ranging from zero in school safety in 2017 to 9.48% in school safety in 2017. Item facet and the interaction effect between school and item facet also explained negligible proportions of the total variance (less than one percent). The correlations between the six subscales were also quite high in the sample of high school parents, with the lowest in 0.79 (clarity of behavioral expectations and fairness of school rules in 2017) and highest in 1.00 between nice pairs of subscales. The high correlations among subscales across years were related to the fact that items in different subscales measure the same aspects of high school parents' perspectives of school climate.

Table 4.26: Estimated Variance and covariance component matrices for high school parents in 2015

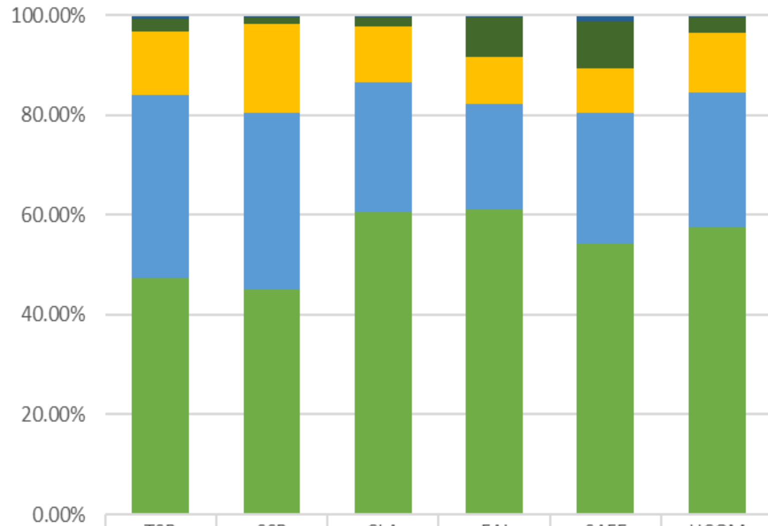
Effect	TSR	SSR	CLA	FAI	SAFE	HCOM
$\hat{\sum} s$						
	0.058	0.985	0.919	0.960	0.966	0.975
	0.066	0.078	0.988	0.979	0.951	0.986
	0.051	0.063	0.053	0.969	0.963	0.987
	0.046	0.054	0.044	0.039	1.000	0.995
	0.049	0.056	0.047	0.042	0.045	1.000
	0.052	0.061	0.050	0.044	0.048	0.050
%	12.7%	17.8%	11.4%	9.2%	9.0%	11.9%
$\hat{\sum} p:s$						
	0.166					
	0.094	0.155				
	0.086	0.104	0.118			
	0.077	0.066	0.079	0.090		
	0.097	0.055	0.070	0.092	0.130	
	0.105	0.087	0.083	0.084	0.117	0.112
%	36.7%	35.5%	25.8%	21.3%	26.2%	27.0%
$\hat{\sum} i$						
	0.011	0.006	0.008	0.033	0.047	0.013
%	2.5%	1.2%	1.6%	7.9%	9.4%	3.0%
$\hat{\sum} si$						
	0.003	0.001	0.002	0.002	0.005	0.001
%	0.6%	0.3%	0.4%	0.3%	1.0%	0.3%
$\hat{\sum} pi:s$						
	0.214	0.196	0.278	0.258	0.268	0.239
%	47.3%	45.0%	60.6%	61.0%	54.2%	57.5%
G	0.845	0.891	0.848	0.835	0.787	0.853
Phi	0.818	0.880	0.823	0.709	0.617	0.809
$\sigma^2(\delta)$	0.011	0.010	0.009	0.008	0.012	0.009
$\sigma^2(\Delta)$	0.013	0.011	0.011	0.016	0.028	0.012

Table 4.27: Estimated Variance and covariance component matrices for high school parents in 2016

Effect	TSR	SSR	CLA	FAI	SAFE	HCOM
$\hat{\sum} s$						
	0.010	0.923	1.000	1.000	0.982	0.959
	0.015	0.026	0.956	1.042	0.950	0.908
	0.010	0.015	0.009	0.954	1.000	0.941
	0.008	0.014	0.008	0.007	1.000	0.900
	0.014	0.023	0.015	0.014	0.022	0.827
	0.008	0.012	0.007	0.006	0.010	0.007
%	2.3%	5.6%	2.5%	1.9%	5.2%	1.4%
$\hat{\sum} p:s$						
	0.251					
	0.206	0.304				
	0.180	0.156	0.233			
	0.196	0.168	0.191	0.224		
	0.209	0.244	0.168	0.188	0.279	
	0.247	0.191	0.173	0.187	0.192	0.274
%	60.4%	65.2%	64.5%	60.5%	68.0%	58.5%
$\hat{\sum} i$						
	0.003	0.001	0.001	0.001	0.000	0.016
%	0.6%	0.1%	0.1%	0.3%	0.0%	3.3%
$\hat{\sum} si$						
	0.001	0.000	0.000	0.000	0.001	0.000
%	0.05%	0.0%	0.0%	00.0%	0.17%	0.0%
$\hat{\sum} pi:s$						
	0.152	0.134	0.118	0.138	0.109	0.172
%	36.5%	28.8%	32.6%	37.1%	26.5%	36.6%
G	0.599	0.777	0.606	0.548	0.744	0.477
Phi	0.578	0.774	0.600	0.535	0.744	0.372
$\sigma^2(\delta)$	0.006	0.008	0.006	0.006	0.007	0.007
$\sigma^2(\Delta)$	0.007	0.008	0.006	0.006	0.007	0.011

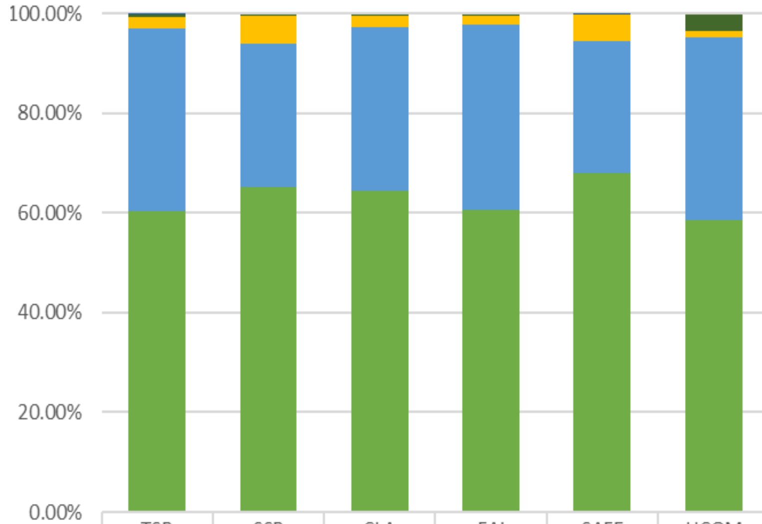
Table 4.28: Estimated Variance and covariance component matrices for high school parents in 2017

Effect	TSR	SSR	CLA	FAI	SAFE	HCOM
$\hat{\sum} s$						
	0.036	0.799	0.983	0.978	0.925	1
	0.026	0.029	0.867	0.792	0.93	0.921
	0.039	0.031	0.043	0.993	1	0.99
	0.04	0.029	0.045	0.047	0.939	0.934
	0.044	0.039	0.052	0.051	0.062	0.983
	0.034	0.028	0.037	0.036	0.043	0.032
%	9.0%	7.1%	11.7%	11.1%	15.5%	6.4%
$\hat{\sum} p:s$						
	0.241					
	0.192	0.287				
	0.191	0.160	0.250			
	0.210	0.170	0.230	0.262		
	0.193	0.216	0.188	0.194	0.237	
	0.254	0.197	0.2	0.222	0.194	0.312
%	60.2%	70.5%	67.8%	61.5%	59.4%	64.0%
$\hat{\sum} i$						
	0.004	0.001	0.001	0.002	0.000	0.01
%	1.1%	0.2%	0.3%	0.5%	0.0%	2.1%
$\hat{\sum} si$						
	0.001	0.001	0.001	0.001	0.002	0.003
%	0.3%	0.2%	0.3%	0.1%	0.7%	0.6%
$\hat{\sum} pi:s$						
	0.117	0.089	0.074	0.114	0.097	0.13
%	29.3%	21.9%	20.1%	26.7%	24.2%	26.7%
G	0.802	0.742	0.829	0.831	0.864	0.723
Phi	0.787	0.741	0.829	0.823	0.864	0.683
$\sigma^2(\delta)$	0.009	0.010	0.009	0.010	0.010	0.012
$\sigma^2(\Delta)$	0.010	0.010	0.009	0.010	0.010	0.015



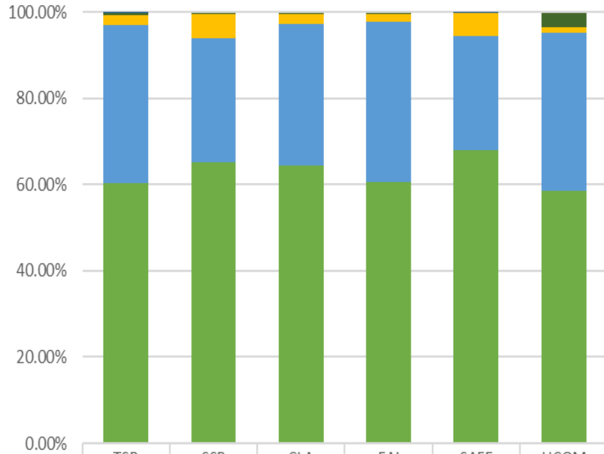
	TSR	SSR	CLA	FAI	SAFE	HCOM
si	0.60%	0.30%	0.40%	0.30%	1.00%	0.30%
i	2.50%	1.20%	1.60%	7.90%	9.40%	3.00%
s	12.70%	17.80%	11.40%	9.20%	9.00%	11.90%
p:s	36.70%	35.50%	25.80%	21.30%	26.20%	27.00%
pi:s	47.30%	45.00%	60.60%	61.00%	54.20%	57.50%

Figure 4.25: Proportions of each variance component to the total variance for high school parents in 2015.



	TSR	SSR	CLA	FAI	SAFE	HCOM
si	0.05%	0.00%	0.00%	0.00%	0.17%	0.00%
i	0.60%	0.10%	0.10%	0.30%	0.00%	3.30%
s	2.30%	5.60%	2.50%	1.90%	5.20%	1.40%
pi:s	36.50%	28.80%	32.60%	37.10%	26.50%	36.60%
p:s	60.40%	65.20%	64.50%	60.50%	68.00%	58.50%

Figure 4.26: Proportions of each variance component to the total variance for high school parents in 2016.



	TSR	SSR	CLA	FAI	SAFE	HCOM
si	0.05%	0.00%	0.00%	0.00%	0.17%	0.00%
i	0.60%	0.10%	0.10%	0.30%	0.00%	3.30%
s	2.30%	5.60%	2.50%	1.90%	5.20%	1.40%
pi:s	36.50%	28.80%	32.60%	37.10%	26.50%	36.60%
p:s	60.40%	65.20%	64.50%	60.50%	68.00%	58.50%

Figure 4.27: Proportions of each variance component to the total variance for high school parents in 2017.

In summary, the G theory estimated variance and covariance components showed similar patterns across respondent groups, grade levels, and years. The respondents within school facet accounted for most of the total variances, which indicates that the DSCS was able to discriminate between individuals' perceptions of school climate reliably. On most subscales, the residual variance (i.e., interaction between school, respondents, items, and other unexplained factors) was the second largest variance component, indicating there was substantial variability in school climate subscale scores attributable to interactions between respondents, schools, items, and other unexamined and unaccounted for sources of variance. The variance of the school facet accounted for around 15% of the total variance, indicating individual's rating of school climate varied to some degree at the school level. Across all samples, item facet and the interaction between school and item facet accounted for a small proportion of the total variances, indicating individual's rating of school climate factors was quite stable across items and the interaction between items and schools.

4.1.4 G and Phi Coefficients of Subscale Scores in Elementary Schools

In the D studies, coefficients and error variances of subscale scores for the current sample were calculated using variance and covariance components for each facet, the number of students within a school, number of items, and variance and covariance component matrices from the G studies. For the elementary school sample, the results of generalizability coefficients (*G*), index of dependability(*Phi*) and error variances are presented from Tables 4.2 to 4.10. Relative error variances ranged from .001 to .004 and absolute error variances ranged from .001 to .036 across grade levels. The results of g and *phi* coefficients are summarized below.

In elementary schools, subscales scores were more reliable for students and teachers than parents, as shown in Figures 4.28. In the student sample, g coefficients of all the subscales scores were larger or close to the .80 threshold considered sufficiently reliable to make decisions about individual differences based on their observed score (Webb, Shavelson, & Haertel, 2006) across the three years. In the teacher sample, g coefficients of five subscale scores were larger or close to the .80 threshold across years. In parents' sample, only the g coefficient of student-student relations subscale score was larger or close to the .80 in 2015 and 2016. Subscale scores were more reliable in student-student relations and school safety than other subscales among all the sample across years. The least reliable subscale scores were clarity of behavioral expectations and fairness of school rules in the student sample, teacher-student relations and teacher-home communication in teacher sample, teacher-home communication in the parent sample. G coefficients across all groups in 2017 were smaller than those from other years.

Figure 4.28: G and Phi coefficients of subscale scores in elementary schools

Regarding the absolute decisions, subscale scores were more accurate in the teacher sample than student and parent sample. G coefficients of subscale scores were larger or close than .80 in five subscales for the teacher sample and one subscale for the student sample across the three years. As indicated by the G studies results, universe score variance (i.e., school variance) was larger in the teacher sample than student and parent samples. Thus, phi coefficients were more robust to the impact of absolute error variance in teacher sample than other groups. The largest phi coefficients of subscale scores were in student-student relations and the smallest coefficients were in fairness of school rules among respondent groups across the three years. Phi coefficients were smaller in 2017 than in other years among the groups.

4.1.5 G and Phi Coefficients of Subscale Scores in Middle Schools

In middle schools, the results of generalizability coefficients (*g*), index of dependability (*phi*) and error variances are presented from Tables 4.11 to 4.19. Relative error variances ranged from .005 to .020, and absolute error variances ranged from .006 to .028 across grade levels. The results of g coefficients and index of dependability are summarized below.

Regarding the relative decisions, subscale scores were more reliable in middle schools than elementary schools across respondent groups and years. In middle schools, g coefficients of subscale scores for the student sample at school level were acceptable in all subscale scores if using the threshold of .80. For the teacher sample, g coefficients of five subscale scores exceeded the threshold of .80 across years, including student-student relations, school-wide engagement, school safety, school-wide bullying, and staff relationships. For the parent sample, the subscale scores of student-student relationships and school safety exceeded the .80 threshold across years. Similar to the results of elementary schools, the subscale scores of student-student relations and school safety tended to be more reliable than other subscales scores across respondent groups and years. The least reliable subscale scores were clarity of expectations for the student sample, teacher-home communication and teacher-student relations for the teacher sample, and teacher-home communication for the parent sample. G coefficients of subscale scores were more stable for the student sample than teacher and parent sample across years. Because of lower school variance in 2016, g coefficients of subscale scores were lower for the teacher and parent sample in 2016.

Regarding the absolute decisions, subscale scores were also more reliable in middle schools than elementary and high schools, as shown in Figures 4.29. In middle schools, phi coefficients of subscales were larger for the student sample and teacher sample than the parent sample. For the middle school student sample, the phi coefficients of four subscale scores exceeded or were close to the threshold of .80 across years, including student-student relations, school-wide engagement, fairness of school rules, and school-wide bullying. For the teacher sample, phi coefficients of three subscale scores exceeded to .80 threshold across years include student-student relations, school safety, and staff-relations. For the parent sample, phi coefficients of student-student relations and school safety subscale scores exceeded to .80 threshold. The subscale scores of student-student relations were more reliable than other subscale scores across respondent groups and years.

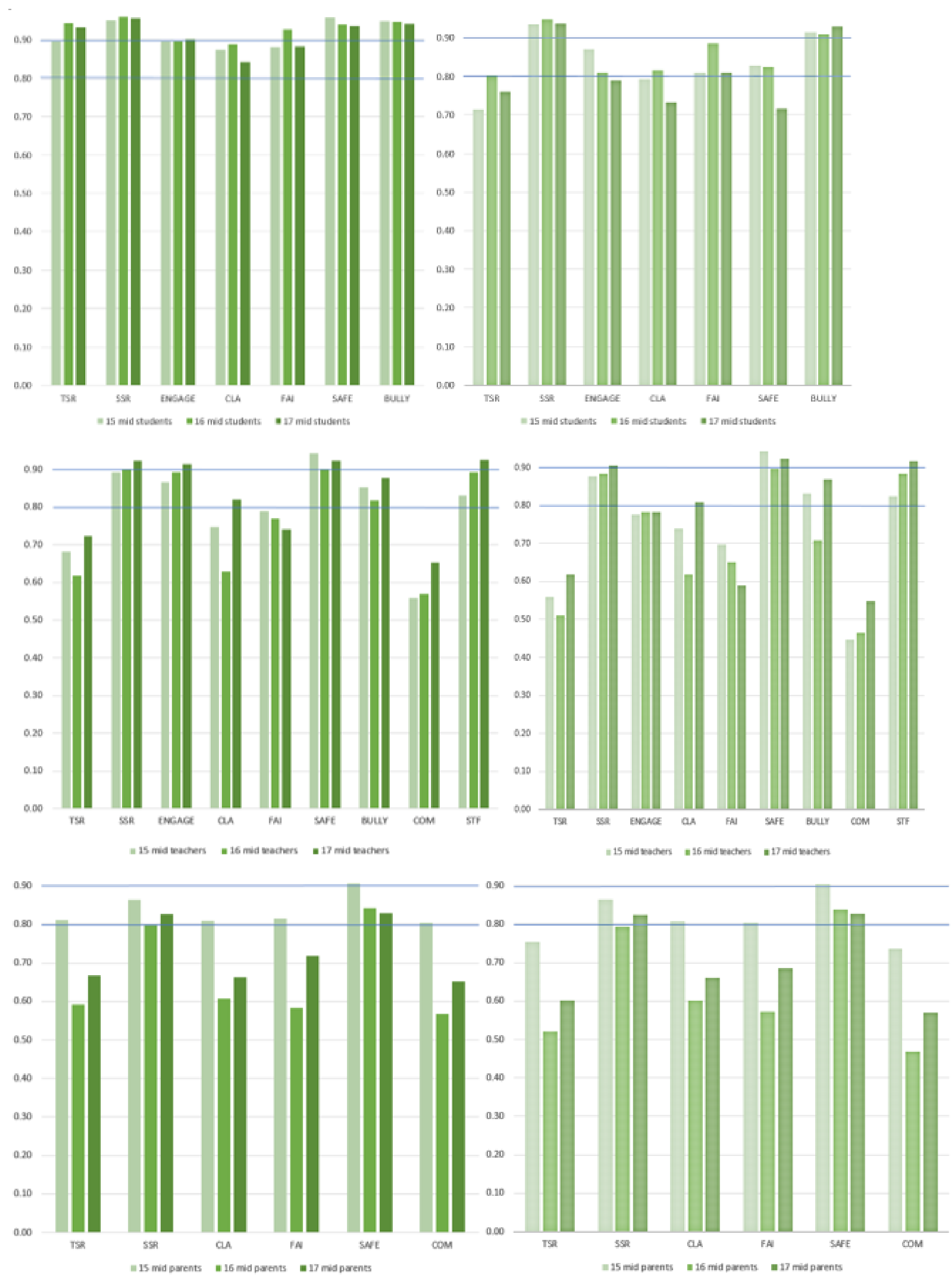


Figure 4.29: G and Phi coefficients of subscale scores in middle schools

4.1.6 G and Phi Coefficients of Subscale Scores in High Schools

The results of *G* coefficients, *Phi* coefficients, and error variances are presented in Tables 4.20 to 4.28. Relative error variances ranged from .002 to .012 and absolute error variances ranged from .003 to .028 across grade levels. The results of G coefficients and index of dependability are summarized below.

In terms of the relative decisions, subscale scores in high schools were more reliable in 2015 and 2017 than 2016 across respondent groups, as shown in Figure 4.30. In high schools, G coefficients of subscale scores for the student sample at school level were acceptable in all subscale scores if using the threshold of .80 in 2015 and 2017. For the teacher sample, G coefficients of six subscale scores exceeded or close to the threshold of .80 in 2015 and 2017, including student-student relations, school-wide engagement, clarity of expectations, school safety, school-wide bullying, and staff relationships. For the parent sample, G coefficients of five out of six subscales exceeded the .80 threshold in 2015 and 2016. G coefficients tended to be higher in school safety subscale across groups and years. The least reliable subscale scores were teacher-student relations in the teacher group and teacher-home communication in the teacher and parent group, especially in 2016. As in middle schools, G coefficients of subscale scores were more stable for the student sample than teacher and parent sample across years in high schools.

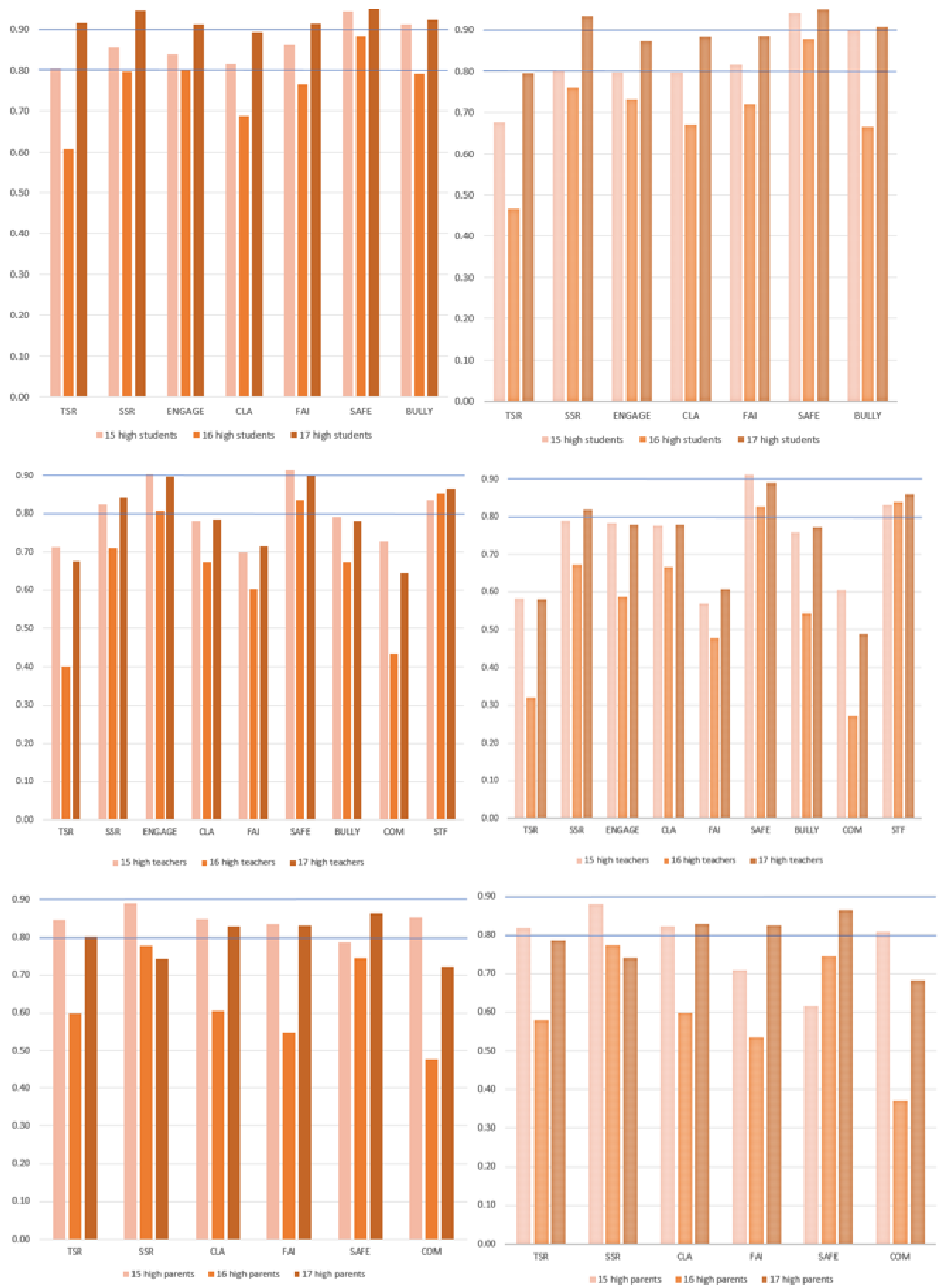


Figure 4.30: G and Phi coefficients of subscale scores in high schools

Regarding the absolute decisions, Phi coefficients of subscales scores were larger in 2015 and 2017 than 2016 across all respondent groups. For the high school student sample, the phi coefficients of the student-student relations and school safety subscale scores exceeded or were close to the threshold of .80 across years. For the teacher sample, Phi coefficients of two subscale scores exceeded to .80 threshold across years including school safety, and staff-relations. For the parent sample, Phi coefficients of student-student relations subscale scores were close to .80 threshold. The subscale scores of student-student relations were more reliable than other subscale scores across respondent groups and years.

4.1.7 Composite G and Phi Coefficients across Groups

Table 4.29 displays the G coefficients, relative error variance, Phi coefficients, absolute error variance of the DSCS scale scores for the 27 groups (3 respondent groups x 3 grade levels x 3 years). Overall, composite G and Phi coefficients were promising and range from .67 in elementary school parent group in 2017 to .97 in middle school student group. Seven out of 27 composite G coefficients exceeded the .80 threshold considered sufficiently generalizable to make relative decisions about schools based on their observed scores. Across years and grade levels, composite G and Phi coefficients were larger in student group than teacher and parent group. Also, composite relative and absolute error variances were smaller in student group than teacher and parent group.

Table 4.29: Composite G and Phi coefficients and error variances for School Climate Scale Scores in MGT model 1

	Students				*Teachers*				*Parents*			
Grade level by year	*G*	$\sigma^2(\delta)$	*Phi*	$\sigma^2(\Delta)$	*G*	$\sigma^2(\delta)$	*Phi*	$\sigma^2(\Delta)$	*G*	$\sigma^2(\delta)$	*Phi*	$\sigma^2(\Delta)$
2017												
Elementary schools	.933	.001	.882	.002	.826	.006	.819	.006	.674	.003	.668	.003
Middle schools	.962	.000	.940	.001	.928	.003	.916	.003	.787	.003	.780	.003
High Schools	.961	.001	.953	.001	.873	.003	.859	.004	.839	.007	.837	.007
2016												
Elementary schools	.944	.001	.910	.002	.890	.006	.885	.006	.777	.002	.770	.002
Middle schools	.972	.001	.957	.001	.906	.004	.893	.002	.743	.003	.735	.003
High Schools	.833	.002	.814	.002	.769	.003	.739	.004	.715	.005	.710	.005
2015												
Elementary schools	.920	.001	.869	.002	.874	.005	.867	.006	.787	.002	.781	.002
Middle schools	.965	.001	.951	.001	.909	.004	.900	.005	.884	.002	.879	.002
High Schools	.920	.001	.910	.001	.886	.004	.872	.005	.913	.005	.902	.006

Note. *G* represents the generalizability coefficient. $\sigma^2(\delta)$ is relative error variance. *Phi* represents the index of dependability. $\sigma^2(\Delta)$ is the absolute error variance.

4.1.8 Variations across Occasions, Respondent Groups, and Grade Levels

The G studies results of MGT model 2 were analyzed separately by grade levels and presented in Table 4.30. In elementary schools (*n* = 48), school facet, the object of measurement, is the largest variance component and explained over 60% of the total variance. The residual facet is the second largest variance component, which accounted for over 19% of the total variance. The occasion variance component explained a negligible proportion of the total variance, which ranged from zero to 2% and indicated that the school climate is stable across measurement occasions. The correlations between respondent groups were high and above .76, which suggested strong agreement between student, teacher, and parent perception of school climate at the school level.

In middle schools (*n* = 12), the residual facet explained the largest proportion of the total variance in student and teacher respondent groups and school facet is the largest variance component in parent respondent group. The variance of occasion facet in student respondent group is twice larger than it's in parent respondent group. However, the occasion variance in teacher group is negative and commonly treated as zero in subsequent calculations. The correlation between students and teachers was .55, and the correlation between students and parents was .79. However, the correlation between teacher and parents at school level were quite small (*r* = .15).

In high schools (*n* = 6), school facet was the largest variance component that accounted for almost 60% of the total variance. In parent respondent group, the school facet covariances are negative and usually treated as zero in mGENOVA calculations. The residual variance is the second largest variance components, ranges from 16% in teachers 100% in parent respondent groups. Due to the negative school covariance in parent group, only the correlation between students and teachers school profile scores was available (r =.49).

The composite G and Phi coefficients of MGT model 2 are presented in Table 4.30, ranging from .96 G coefficient in elementary schools and to .59 Phi coefficient in high schools. The variances of occasions were relatively small across grade levels so the differences between G and Phi coefficients were minimal. In longitudinal assessment, school climate profile scores were more reliable and precise in elementary schools than middle and high schools.

Table 4.30: G and D studies results for MGT model 2

	Elementary schools			Middle schools			High schools		
Effect	Students	Teachers	Parents	Students	Teachers	Parents	Students	Teachers	Parents
$\hat{\Sigma}_s$									
	0.010	0.774	0.795	0.003	0.545	0.787	0.008	0.497	0.000
	0.013	0.027	0.768	0.002	0.007	0.146	0.006	0.021	0.000
	0.005	0.007	0.003	0.003	0.001	0.006	0	0.011	0
%	73.93%	80.85%	62.60%	45.83%	35.02%	61.89%	59.15%	62.68%	0.00%
$\hat{\Sigma}_o$									
	0.000			0.001			0.000		
	0.000	0.000		0.000	0.000		0.000	0.007	
	0.000	0.000	0.000	0.000	0	0.000	0	0.004	0.000
%	1.79%	0.00%	2.00%	9.00%	0.00%	4.33%	0.51%	21.30%	0.00%
$\hat{\Sigma}_{so}$									
	0.003			0.003			0.006		
	0.001	0.006		0.002	0.013		0	0.005	
	0.001	0.000	0.002	0.001	0.004	0.003	0.006	0	0.073
%	22.79%	19.42%	31.00%	47.50%	64.98%	33.22%	40.35%	16.01%	100.00%
G	0.928	0.943	0.890	0.744	0.618	0.848	0.746	0.887	0.000
Phi	0.923	0.943	0.883	0.709	0.618	0.832	0.743	0.771	0.000
Composite G coefficient		0.963			0.725			0.624	
Composite Phi coefficient		0.963			0.738			0.599	

Note. The sample sizes of elementary, middle, and high were 45, 12, and 6.

4.2 Optimization of Measurement Procedures

The G and Phi coefficients of the school-level scale scores were above the threshold of .80 considered as a sufficient reliable index to make measurement decisions in most groups, as shown in Table 4.29. If decisions for future measurement procedures focus on the school-level scale scores, reducing the sample size per school might be a strategy to reduce the cost of survey administration and response burden, especially in student and teacher respondent groups. At the same time, this strategy can also maintain the reliability of scale scores at the appropriate threshold. However, if the decision concentrates on subscale scores, increasing the sample size per school and the number of items might be an option to improve the quality of measurement precision. To satisfy both types of possible measurement decisions, it is necessary to have a spectrum of manipulation conditions and provide a cost-effect solution to meet the considerable threshold of .80.

4.2.1 Optimizations in MGT Model 1

The effects of manipulating the levels of the random facets in a series of D studies were examined. Because the results of G studies showed similar patterns across years, only the 2017-year data were used to calculate both G and Phi coefficients across groups by changing the number of respondents within schools (i.e., 40%, 60%, 80%, 100%, 120% of original respondents per school) and the number of items per subscale (i.e., decreasing one item per subscale, original design, and increasing one item per subscale). In this phase, 135 D studies were conducted. It would require many pages to describe the results of every single D study. Thus, the respondent groups with highest and lowest composite G and Phi coefficients were chosen as representations of the D studies results. The results of the rest D studies were also summarized.

Middle school students group has the highest G and Phi coefficients for the DSCS scale scores. Figure 4.31 illustrates the effects of manipulating the number of students per school and the number of items per subscale on the composite G and Phi coefficients for scale scores.

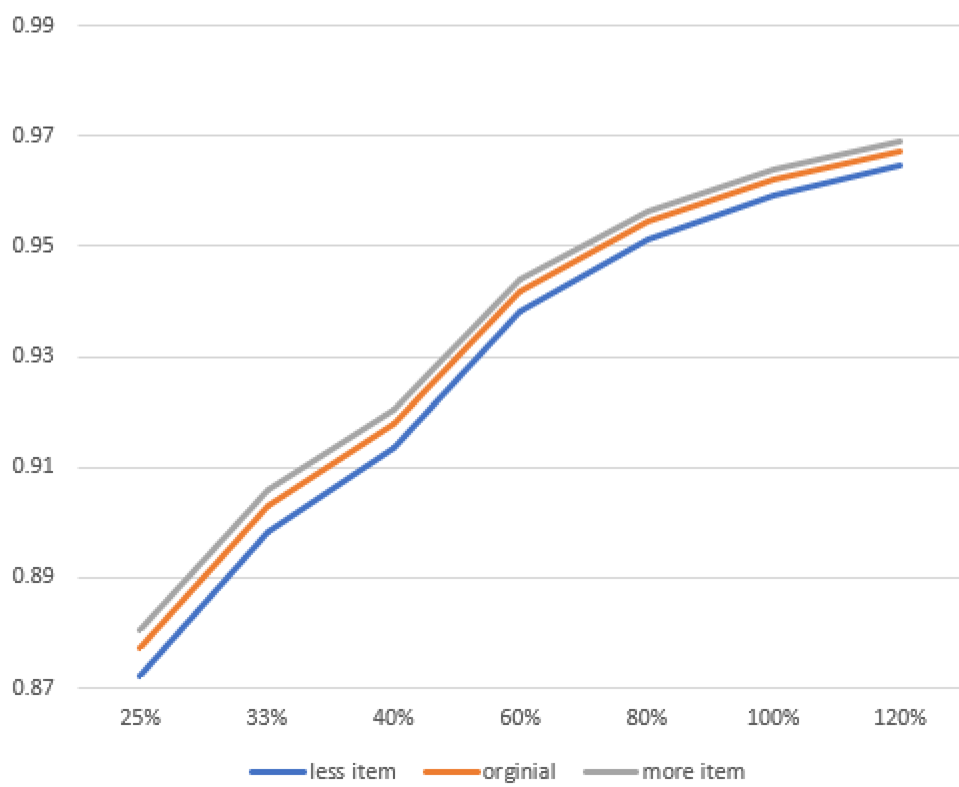


Figure 4.31: A serious of optimization D studies for the middle school students sample

Even with the most economical option, the composite G and Phi coefficients were above the threshold of .80 to achieve a reliable estimate of school-level scale scores and the .90 threshold for critical decisions when using school-level scale scores. If the measurement decision focuses on the relative evaluation of subscale scores, 80% of the original number of students per schools was enough to reach the .80 threshold across all subscales. The use of the DSCS subscale scores is a relative decision to compare schools in the State of Delaware, but there is a possibility that the subscales scores might be used to compare schools at the population level, nationwide. Under that scenario, the most "expensive" measurement procedure in this dissertation cannot reach the threshold of .80 across subscales. For teacher-student relations, clarity of expectations, and school safety subscales, the phi coefficients were close to .80.

The elementary school parents group has the lowest composite G and Phi coefficients of scale scores. None of the proposed measurement procedures exceeded the .80 G and Phi coefficients thresholds given the current parent sample sizes. Although the DSCS survey administration procedures require 100% of the students' parents or guardians to answer the survey and return it in an envelope to ensure parents' confidentiality, response rates for parents are low. Therefore, it may be possible to increase response rates somewhat, but it may not be realistic to increase the parent sample by 200%.

As indicated by the G study results, the school level variance accounted for around 2% of the total variance across subscales. The variance at the individual level accounted for around 70% of the total variance in parents' groups across grade levels. It is possible that individual parents' perceptions of school climate vary drastically different within a school, but these individual differences might be "averaged out" at the school level.

The results of the additional D studies indicated that G and Phi coefficients of scale and subscale scores increased as the number of respondents per school and items per subscale increased. If the measurement decisions are focused on the scale scores, fewer respondents per school and fewer items were needed to reach .80 G and Phi coefficients in almost all the groups, except for the 2017 elementary and middle school parent groups. In all the student groups and middle school teacher groups, the results of D studies indicated that administering the survey to about half of the respondents per school and one item less per subscale were enough to reach the threshold of .80 if the measurement decisions concerns about the school-level scale scores. In the elementary teacher group, the same respondents per school and one less item per subscale might enough to reach the threshold of .80 to provide reliable estimates of school-level scale scores. In middle school parent group, 120% of the original sample size for parents per school might be enough to reach the threshold of .80. In the high school parent group, 80% of the original sample of parents per school might be enough to reach the threshold of .80.

The D studies results of G and Phi coefficients of subscale scores are not as robust as those for scale scores. The G coefficients of subscale scores exceed the threshold of .80 in middle school student group with 80% of students per school and high school student group with 60% students per school and one less item per subscale. In the rest of the groups, even with the most “expensive” optimization procedure, the G and Phi coefficients of subscale scores still could not exceed the threshold of .80 across subscales because the residual variance of those subscales accounted for a large proposal of the total variance.

4.2.2 Optimization in MGT Model 2

The purpose of MGT model 2 is to investigate the generalizability of the DSCS scores in longitudinal assessment. The school profiles produced by MGT model 1 was the dependent variable in model 2. School is the object of measurement, occasions is a random facet, and respondent groups is the linked facets. Figure 4.32 showed an increase in G and Phi coefficients as the number of years increases from two to five. Patterns were consistent in elementary, middle, and high schools. The differences between G and Phi coefficients were relatively small as the occasion variances were small, which are the only difference between those two coefficients.

Figure 4.5 illustrates that at least two years are required to reach .80 and above G coefficients and at least three years are needed to obtain .80 and above Phi coefficients for student, teacher, parent subscores, and school profile scale scores in elementary schools. It also shows that at least five years are required to reach .80 and above G and Phi coefficients for student, parent subscores, and school profile scale scores in middle schools. The figure also shows that at least five years are required to reach .80 and above G coefficients and at least six years are needed to obtain .80 and above Phi coefficients for student and teacher subscores, and school profile scale scores in high schools.

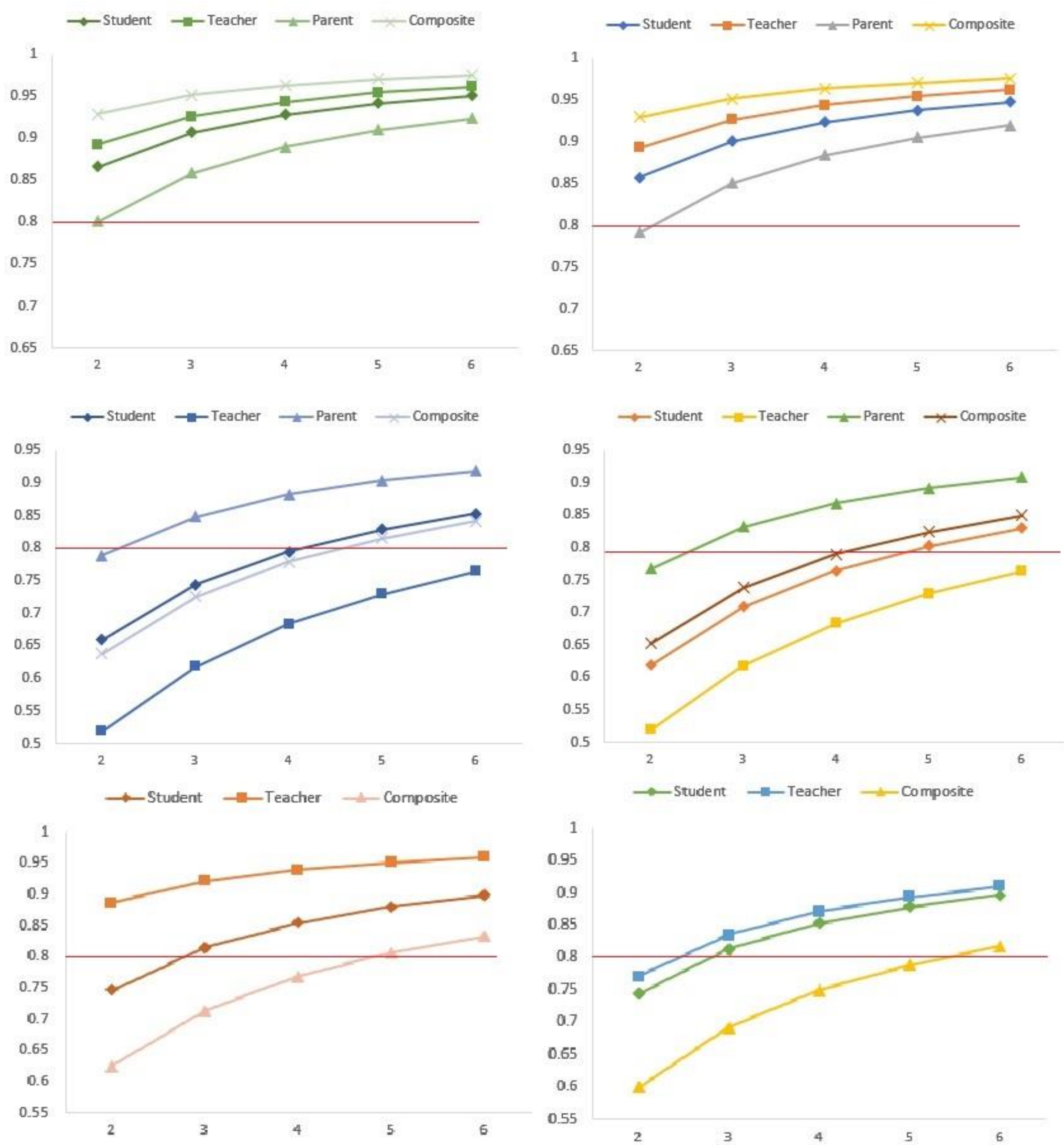


Figure 4.32: G and Phi coefficients in MGT model 2 optimization procedures. In the first row, the left-hand chart shows the change of g coefficients with increasing the number of years in elementary schools. The right-hand chart shows the change of Phi coefficients when adding more measurement oaccasions in the elementary schools. The second and third rows, repectively, show these patterns in middle and high schools.

Chapter 5

DISCUSSION

5.1 Advantages of Applying Multivariate Generalizability Theory to the DSCS

The results of this dissertation contribute to the conceptual and empirical understandings of school climate as a multilevel, multi-informant (i.e., student, teacher, parent), and multi-dimensional construct and how to accurately measure school climate considering various sources of error using the DSCS. Empirically, multivariate generalizability theory (MGT) partitions variation in scale scores into components related to schools, respondents within schools, items, subscales, and occasions. By determining which sources accounted for the greatest proportion of total variance, procedures can be designed to reduce those sources of error and optimize the efficiency and precision of the DSCS scale and subscale scores. Conceptually, MGT can help achieve better understandings of factors related to measuring school climate and the relationships between those factors.

The multi-dimensional aspect of measuring school climate can be treated as the linked facet in MGT to investigate the relative contributions of subscales to the reliability of school climate scale scores and relationships between these subscales. The multilevel aspect, school, can be considered as the object of measurement, and the respondents within schools a facet. This aspect confirms that the DSCS school climate scores can be used as school-level indices to reliably distinguish schools and make inferences about the schools.

There are two ways to handle the multi-informant aspect of measuring school climate: the first MGT model treats respondent groups separately and provides detailed diagnostic information; the second MGT model adds the multi-informant component as the linked facet to explore the relative contribution of each respondent group to the longitudinal assessment of school climate and homogeneity between respondent groups in a school. The temporal issue of measuring school climate can be considered as a random facet, occasion, in the second MGT model, which examines the variation of occasion and the interaction between occasion and schools.

5.2 Quality of the DSCS Measurement Precision

The first research question asked what multivariate generalizability theory tells us about the measurement design and precision of the Delaware School Climate Scale, including: (1) which sources of variation contribute to the school climate scale and subscale scores; (2) which types of respondents' ratings of school climate are more reliable across occasions. The G studies estimates were calculated based on the variance and covariance components of all the facets in the two models. In MGT model 1, school was the object of measurement, individual response was the dependent variable, subscale was the linked facet, and person within school and item within subscale were the two random facets. The analysis of MGT model 1 was conducted separately by respondent groups, grade levels, and occasions. In MGT model 2, school was also the object of measurement, school profile scores obtained from model 1 formed the dependent variable, the respondent group was the linked facet, and occasion was a random facet. The analysis of MGT model 2 was conducted separately by grade levels.

For MGT model 1, results of G studies were relatively consistent across grade levels, respondent groups, and occasions. The two largest variance components were person with school variance and residual variance (i.e., the interaction between items, schools, and respondents, and other unknown sources of error). These two components accounted for almost 80% of the total variance across all the groups. Between 1% to 33% of the total variance can be attributed to the school, which is the desired variance of the object of measurement. Item variance and the interaction between items and school variance accounted for a minimal proportion of the total variance. These findings further support the idea that some aspects of school climate (i.e., student-student relations, school-wide bullying, school safety, and teacher-staff relations, especially in teacher group) can be treated as shared school-level perceptions of school climate, whereas other aspects (i.e., teacher-student relations, clarity of behavioral expectations, fairness of school rules) should be considered as multilevel observations with considerable variance across individuals' perceptions.

5.2.1 Between and within School Variances

Several empirical studies have used school climate scores as an outcome measure for evaluating the impact of school interventions or as a predictor for policy-relevant evaluations (Shear et al., 2008). If using school climate scores to evaluate schools, the average scores of individuals in the school is the variable of interest, which is also called the object of measurement in MGT models. This dissertation found that school variations in different subscales accounted for a certain proportion (ranging from 1% to 33%) of the total variance. On average, school variances of subscale scores were larger in the teacher group than the parent and student groups. It is concerning that school variances for some scales accounted for a minimal proportion of total variance, as the purpose of the DSCS is to measure school climate at the aggregated unit, school. However, this finding is consistent with prior research. Other studies have reported that school-level school climate accounts for less than a quarter of the total variance in measuring school climate (Schweig, 2014) or other similar school-level variables (Beem, Brugman, H\ost, & Tavecchio, 2004).

One possible explanation for low school variance is the lack of relative references in measuring the organizational environment, especially in the student and parent groups. It is assumed that the climates of different schools can be measured and distinguished by aggregating individual responses to the school level (e.g., by averaging). However, it is challenging for individual respondents to objectively answer questions that aim to measure a psychological construct of a group of which they are a part. First, respondents might have limited knowledge of what the school climate looks like in other schools, especially for the student and parent groups. Without enough information about other reference schools, individual evaluations of schools largely depend on their experiences in their own schools. It may also explain why there were larger school variances for the teacher respondent group than the parent and student groups. Teachers have more opportunities (e.g., professional development sessions, social gathering in the district) and access to the information related to school climate in other schools. Students and parents are more likely to stay in the same neighborhood over the years and less likely to be attuned to what happens in other schools.

Also, even though the individual DSCS items were intended to measure school characteristics, respondents may respond to the items based on several factors related to their own experiences, such as recent events at school, personality, and academic achievement. These factors might be more powerful influences on respondents' ratings of school climate rather than objective evaluations of school climate. As presented in the results, the person within school facet accounted for a greater proportion of total variance than the school facet in all the groups across all years. This indicates that some aspects of school level school climate perceptions contain more "noise" (within school variation) than "signal" (between school variation). The present findings seem consistent with other studies with similar analytic models that found that individuals differences (i.e., within school variation) accounted for the majority of the total variance in measuring school climate or school moral atmosphere (Beem et al., 2004; Konold et al., 2014; Schweig, 2014).

5.2.2 Item and Item by School Variances

The results of the first MGT model found that the item variances were minimal or even negligible across all the groups, which indicates that respondents' rating of school climate items were homogeneous within each subscale. It might be no surprise that the variance of the item by school interaction was also relatively small, meaning that ratings of school level DSCS items were quite stable. These findings were consistent with those of other studies (e.g., Scheweig, 2014) and implied that the DSCS items within each subscale were consistent and those subscales can be used for making reliable cross-school comparisons. Furthermore, the small item by school interaction suggests that the scale displays the desirable property of measurement invariance across schools (i.e., the measurement constructs are consistent across schools).

5.2.3 Occasion Variation across Grade Levels

The purpose of the second MGT model was to investigate the contribution of occasions when using multi-informant school climate scores (i.e., student, teacher, parent) longitudinally in elementary, middle, and high schools. Using school profile scores in longitudinal assessment, results of G studies showed different patterns across grade levels. In elementary schools, the distribution of variance components (i.e., school, occasion, school X occasion) was consistent across the students, teachers, and parents respondent groups, which indicates that the relative contribution of each facet to the student, teachers, and parent ratings of school climate from 2015 to 2017 was essentially the same. The results were less encouraging in middle and high schools as the distribution of variance components varied across respondent groups. These findings indicate that perceptions of school-level school climate over the years are more homogeneous between respondents in elementary schools than in middle and high schools. Divergent perceptions of school climate may imply disagreement between groups or even potential conflicts in schools.

Across grade levels, the occasion facet explained a minimal proportion of total variance, except in the high school teacher group. These findings are a desirable psychometric property of the DSCS, a stable measurement structure of the scale was expected across years. Also, these findings further support the idea of viewing school climate as an organizational phenomenon. Changes in organizational environments take time, and such is the case with school climate; consistency across a five-year span is not surprising. Another important finding of MGT model 2 is the relatively substantial contribution of the interaction between occasion and schools to the total variance, which indicates the changes in school climate over time varies across schools. This finding is an essential implication to the longitudinal evaluation of school climate as researchers are now able to reliably examine the change in school climate over time using school profile scores.

5.2.4 G and Phi Coefficients

The use of school climate instruments, in most instances, is intended to distinguish perceptions of school climate between schools and make inferences about schools. Previous validation studies of school climate measurement (including the DSCS) only reported the reliability coefficients of school climate at the individual rater level, which does not align with the use of school climate scores at the school level. This dissertation is the first study using multivariate generalizability theory to evaluate the reliability of scores in measuring school-level school climate simultaneously considering multiple sources of error.

In the current measurement design, the composite G and Phi coefficients of the DSCS scale scores were promising. In MGT model 1, the results of D studies indicated that the DSCS yields scale scores that can support reliable relative and absolute inferences regarding students' and teachers' perceptions of school climate at a considerable reliability criterion (≥ .80; Webb et al., 2006). Regarding students' perceptions, the scale scores can support such inferences even at a stringent threshold recommend for high-stakes decisions (≥.90; Nunnally, 1976). The composite G and Phi coefficients of parents' ratings were close to the considerable threshold of .80. Across samples, the subscale scores were less reliable than scale scores, as indicated by smaller G and Phi coefficients and larger error variances. Only in middle schools, did the coefficients of all subscales scores exceed the .80 threshold for low-stakes decisions. These findings indicated that the DSCS can provide reliable and precise estimates of students' and teachers' perceptions of school-level school climate across grade levels and respondent groups. In MGT model 2, the results of D studies showed school profile scores might be considered as reliable measures to distinguish schools among respondent groups across years in elementary schools. Although the G and Phi coefficients were relatively lower in middle and high schools, it should be noted that the limited number of schools at these grade levels might skew the results. Thus, the lower coefficients should be intercepted with caution in middle and high schools.

5.3 Optimization of Future Measurement Procedures

The second research question asks about how multivariate generalizability theory can optimize the reliability of the DSCS scores when simultaneously considering various sources of variance. If the future measurement procedure focuses on students, fewer students (40% of the original number of students per school) and one less item per subscale is necessary to reach the recommended threshold .80 across grade levels. For the teachers' sample, fewer teachers per school (60% of original number of teachers per school) are needed to achieve the threshold of reliability in middle and high schools. For the parents' sample, fewer parents per school (80% of the original number of parents per school) are required to achieve the recommended threshold in high schools, but more parents per school are required in elementary and middle school. However, regarding subscale scores, even the most "expensive" measurement design (i.e., 120% of the original person per school and one more item per subscale) could not reach the threshold of .80 if the measurement procedures focus on the relative and absolute inferences of subscale scores across respondent groups and grade levels.

5.4 Specific Recommendations for the DSCS Future Measurement Procedures

The findings and arguments from section 5.1 to section 5.3 give rise to the following recommendations for producing the reliable and high-quality measures of school level school climate using the DSCS:

1. Improve elementary and middle school parents' participation in survey administrations so that the number of parents per school can be increased by at least 20%.
2. Consider a sampling plan for future measurement procedures in which the Delaware School Climate Scale is administered to half of the students (randomly selected) in a school and other scales on the Delaware School Surveys are administered to the other half of the students in the school. The length of Delaware School Surveys currently exceeds 100 items. This sampling plan would maintain reliability of the DSCS scores, while cutting response burden in half.
3. Another possible sampling plan for future measurement procedures is to use matrix sampling: splitting the Delaware School Surveys into two or three different forms and randomly assigning respondents to answer one of the survey forms. A $(p^{\circ} : s^{\bullet}) \times i^{\circ}$ multivariate G theory design might apply to the situation where students nested within different schools respond to different sets of items. This might also allow adding one or two additional items to subscales with lower reliability; thus, increasing school-level reliability with more items, but decreasing individual response burden by randomly assigning students to two or three different forms containing a subset of items from each scale.
4. In the teacher scale, it is possible to decrease the number of items per subscale if the research goal focuses on using scale scores.

5. Administer Delaware School Surveys every other year as the change of school climate takes place slowly.
6. Require students, teachers, and parents in a school to respond to the survey within the same administration year to achieve more homogeneous evaluations of school climate, especially in middle and high schools.
7. In middle and high schools, it may be necessary to have at least five years of survey administration to reach reliable school profile scores.

5.5 Implications for Using the DSCS Scale and Subscale Scores

Multivariate generalizability theory offers a rich theoretical framework and a robust set of analytic tools for investigating issues of reliability and precision of surveys (Brennan, 2001). The measurement procedures of the DSCS focus on providing scale and subscale scores of the aggregated units, schools. The class-means design in multivariate generalizability theory has the capacity to deal with the nested structure of the DSCS data and to provide accurate estimations of scale and subscale scores. As discussed in the introduction, the school climate score is often considered as a measurable outcome for illustrating the improvement of school conditions and examining the effectiveness of behavioral interventions. The findings of this dissertation suggest that school climate school profile scores produced by the MGT scaling model can be used as reliable measures to track the change of school climate between schools over several years.

Regarding the use of the DSCS subscale scores, the findings of this dissertation suggest that researchers should use them with caution. The G studies results indicated that school level DSCS subscale scores were highly correlated, which may raise some concerns about the psychometric evidence supporting multi-dimensional constructs from the DSCS. Also, if using the DSCS subscale scores as independent variables in regression models, there is a high likelihood of multi-collinearity between subscales. Finally, the D studies results suggest that subscale scores were less reliable than scale scores across groups, grade levels, and years. The broader literature on the use of subscales scores suggests that the derivation of subscales scores may not provide distinct information that is not already included in the total test scores (Haberman, 2008). A variety of psychometric methodologies have been developed to produce diagnostic information regarding the usefulness of subscale scores (e.g., reliability index for score profiles; Jiang & Raymond, 2018) over the last ten years. Thus, researchers should not overinterpret the meaning of subscale scores, especially when using subscale scores for evaluation purposes regarding behavioral interventions.

5.6 Implications for Measuring School Level School Climate

These findings of G studies have important implications for designing school climate measures and developing measurement procedures for producing school climate scores for individual schools. As discussed, the person within school facet (i.e. within school variation) is commonly found as the largest sources of error variance (Beem, et.al., 2014; Scheweig, 2014) rather than the desired object of measurement, school. To obtain reliable estimations of school-level school climate, it is necessary to have large within-school sample sizes (e.g. many students, teachers, and parents per school) to wash out the effect of "noise." The harmonic mean of school sizes is the denominator in the formula for calculating error variance. If the between school variation is limited, large school sizes can help reduce the impact of within-school variation to the precision of school-level school climate scale and subscale scores. For instance, in this dissertation, as between school variations were limited in both middle school student and parent groups, the G and Phi coefficients of scale scores were more promising in students’ groups than those in parents’ groups as the harmonic means of students per school are almost six times larger than those of parents per school. This finding may help researchers and educators understand how to design appropriate sampling plans to achieve reliable estimates of school-level school climate. Also, the results of within and between school variation provide empirical evidence to support the idea that the focal unit of analyzing school climate should be considered at both the individual and school levels. Neglecting either level of data may cause inconsistent interpretations of school climate and its relationship with other prominent academic and behavioral outcomes.

5.7 Implications for Choosing Informant in Measuring School Climate

As stated in the methodology section, administering the survey to students, teachers, and parents in a school is an expensive and time-consuming process. Consistency is the key to evaluating the homogeneity of students’, teachers’, and parents’ perceptions of school climate longitudinally. In MGT model 2, less than 50% of the original schools – 45 elementary schools, 12 middle schools, and six high schools – successfully executed a consistent survey administration from 2015 to 2017. Although the sample size of schools does not preclude a G theory analysis (Brennan, 2010; Briesch, Swaminathan, Welsh, & Chafouleas, 2014), the numbers of middle and high schools might not be enough to make statistical inferences in longitudinal modeling.

On the other hand, it is understandable that researchers or schools may not have the resources to implement such expensive procedures to collect data from all three informants. The findings discussed in the previous section also offer support for choosing the appropriate informant for assessing accurate school-level school climate. Students and teachers might be the more appropriate groups than parent groups for evaluating school-level school climate for a couple of reasons. First, the G and Phi coefficients of school climate scale and subscale scores by teachers' and students' groups were much larger than those in the parent groups, which indicates that it is more likely to find "true score" differences between schools using scores based on teacher and student responses as opposed to the parent group. In addition, teachers and students may have better knowledge than parents of the social and physical environment of schools. Thus, teachers and students may be more likely to have a consistent and informed evaluation of their schools as they are the main personnel components of the school.

On the other hand, researchers may have concerns about teacher bias in reporting outcome measures related to school-level interventions as teachers usually implement the programs in schools. A student response survey is the most widely used in comparison to teacher and parent responses in school-effectiveness and school reform research (Bear, et al., 2011). Teacher-report bias should be considered as a shortcoming in program evaluations. It should also be noted that the implication for choosing teachers as the most appropriate informant group is based on the results of this dissertation that aiming at measuring school climate using the most recent version of DSCS-S, DSCS-T, and DSCS-H.

Previous empirical studies using the DSCS found that teachers tend to have more positive perceptions of teacher-student relations and student-student relations than students (Bear, et al., 2014). Yet, the sample in the study was composed of five elementary, five middle, and five high schools that with high students' and teachers' response rates, which might fully reflect the variability of school climate scores in the State of Delaware. Instead of focusing on raw differences of relation variables between students and teachers, this dissertation concerns about the variabilities of school climate scores among respondent groups at school-level and individual level.

Alternatively, it is possible that researchers or investigators are only interested in parent perspectives. When measuring parents' perceptions of school level school climate, it is typical to observe minimal variation between schools and substantial variation within schools. For the parent group, researchers should obtain responses from at least half of the parents within a school, especially in elementary schools In the current DSCS sampling plan, the survey is administered to all the students' parents in a school. However, results from the D studies showed that even 120% of the current number of parents per school is insufficient for making relative decisions with the DSCS scale scores.

5.8 Limitations

Results of this dissertation must be interpreted in light of the following limitations. First, hidden facets may exist in the universe of admissible observations beyond those defined this dissertation. For example, the format of administering the Delaware School Surveys may influence the reliability of the DSCS scale and subscale scores. The survey administration of the DSCS allows students and teachers to answer the survey either using a paper survey or an anonymous online survey. It is possible that students with less exposure to technology are less likely to feel comfortable answering survey questions on computers. Thus, hidden facets could bias the estimations of variance and covariance components. The facets included in this dissertation are the most critical factors influencing the precision of the DSCS scale and subscale scores but adding the format of survey administration in a G theory model might be a direction for future research in examining the quality of data collection procedures.

Another limitation of this dissertation concerns the method for handling missing data. Due to the algorithm of mGENOVA software, missing data was not allowed in data analysis, and listwise deletion was used to exclude any respondent who missed one or more questions. Some may argue for using multiple imputations or a different software package to deal with missing data. However, there are no minimum requirements regarding the adequacy of sample size in measurement procedures when conducting G studies or D studies. Also, the computation capacity of mGENOVA is the most efficient among the available packages when estimating generalizability theory models. In other words, the analytic strategies used in this dissertation reflect some compromises. Future advances in software and computational capacity may make it possible to conduct analyses of the DSCS data with a fully-specified MGT model, including item-level missing data.

REFERENCES

Allen, M. J., & Yen, W. M. (2001). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.

Anderson, C. S. (1982). The search for school climate: A review of the research. *Review of Educational Research, 52*, 368-420. doi:10.2307/1170423

Bandyopadhyay, S., Cornell, D. G., & Konold, T. R. (2009). Validity of three school climate scales to assess bullying, aggressive attitudes, and help seeking. *School Psychology Review, 38*, 338-355.

Baumrind, D. (1966). Effects of authoritative parental control on child behavior. *Child Development*, *37*(4), 887-907. doi:10.2307/1126611

Baumrind, D. (1996). The discipline controversy revisited. Family Relations: An Interdisciplinary. *Journal of Applied Family Studies*, 45, 405-414. doi:10.2307/585170

Bear, G. G. (2010). *School discipline and self-discipline: A practical guide to promoting prosocial student behavior.* Guilford Press.

Bear, G. G., Gaskins, C., Blank, J., & Chen, F. F. (2011). Delaware school climate Survey—Student: Its factor structure, concurrent validity, and reliability. *Journal of School Psychology, 49,* 157-174. doi:10.1016/j.jsp.2011.01.001

Bear, G. G., Yang, C., Harris, A., Mantz, L. S., Hearn, S., & Boyer, D. (2016). *Technical manual for the Delaware school survey: Scales of school climate; bullying victimization; student engagement; positive, punitive, and social emotional learning techniques; and social and emotional competencies. Delaware Positive Behavior Support (DE-PBS) and School Climate Transformation Projects*

Bear, G. G., Yang, C., & Pasipanodya, E. (2015). Assessing school climate: Validation of a brief measure of the perceptions of parents. *Journal of Psychoeducational Assessment*, *33*, 115-129. doi:10.1177/0734282914545748

Bear, G. G., Yang, C., Pell, M., & Gaskins, C. (2014). Validation of a brief measure of teachers' perceptions of school climate: Relations to student achievement and suspensions. *Learning Environments Research, 17,* 339-354. doi:10.1007/s10984-014-9162-1

Bear, G. G., Yang, C., Glutting, J., Huang, X., He, X., Zhang, W., & Chen, D. (2014). Understanding Teacher-Student Relationships, Student-Student Relationships, and Conduct Problems in China and the United States. *International Journal of School & Educational Psychology*, *2,* 247–260. doi: 10.1080/21683603.2014.883342

Beem, L., Brugman, D., Host, K., & Tavecchio, L. W. (2004). Students' perception of school moral atmosphere: From moral culture to social competence. A generalizability study. *European Journal of Developmental Psychology, 1*, 171-192. doi:10.1080/17405620444000076

Berg, J. K., & Cornell, D. (2016). Authoritative school climate, aggression toward teachers, and teacher distress in middle school. *School Psychology Quarterly, 31*, 122-139. doi:10.1037/spq0000132

Bloch, R., & Norman, G. (2012). Generalizability theory for the perplexed: A practical introduction and guide: AMEE guide no. 68. *Medical Teacher, 34*, 960-992. doi:10.3109/0142159X.2012.703791

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions, 12,* 133-148. doi: 10.1177/1098300709334798

Bradshaw, C. P., Waasdorp, T. E., Debnam, K. J., & Johnson, S. L. (2014). Measuring school climate in high schools: A focus on safety, engagement, and the environment. *Journal of School Health, 84*, 593-604. doi: 10.1111/josh.12186

Brand, S., Felner, R., Seitsinger, A., Burns, A., & Bolton, N. (2008). A large scale study of the assessment of the social environment of middle and secondary schools: The validity and utility of teachers' ratings of school climate, cultural pluralism, and safety problems for understanding school effects and school improvement. *Journal of School Psychology*, *46*, 507-535. doi:10.1016/j.jsp.2007.12.001

Brand, S., Felner, R., Shim, M., Seitsinger, A., & Dumas, T. (2003). Middle school improvement and reform: Development and validation of a school-level assessment of climate, cultural pluralism, and school safety. *Journal of educational psychology, 95*, 570. doi: 10.1037/0022-0663.95.3.570

Brennan, R. L. (2001). *Generalizability theory*. New York, NY: Springer.

Briesch, A. M., Chafouleas, S. M., & Riley-Tillman, T. C. (2010). Generalizability and dependability of behavior assessment methods to estimate academic engagement: A comparison of systematic direct observation and direct behavior rating. *School Psychology Review, 39*(3), 408-421.

Briesch, A. M., Swaminathan, H., Welsh, M., & Chafouleas, S. M. (2014). Generalizability theory: A practical guide to study design, implementation, and interpretation. *Journal of School Psychology, 52,* 13-35. doi:10.1016/j.jsp.2013.11.008

Bronfenbrenner, U. (1975). Reality and research in the ecology of human development. *Proceedings of the American Philosophical Society*, *119*(6), 439-469.

Bronfenbrenner, U., & Morris, P. A. (2006). *The bioecological model of human development. Handbook of child psychology (6th ed.): Vol 1, theoretical models of human development (pp. 793-828)*. Hoboken, NJ, US: John Wiley & Sons Inc.

Brophy, J. E. (1996). *Teaching problem students*. New York, NY: Guilford Press.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Publications.

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

Buhs, E. S., Ladd, G. W., & Herald, S. L. (2006). Peer exclusion and victimization: Processes that mediate the relation between peer group rejection and children's classroom engagement and achievement? *Journal of Educational Psychology, 98*, 1-13. doi:10.1037/0022-0663.98.1.1

Cardinet, J., Johnson, S., & Pini, G. (2010a). *Applying generalizability theory using EduG*. New York, NY: Routledge/Taylor & Francis Group.

Centers for Disease Control and Prevention. (2009). *School connectedness: Strategies for increasing protective factors among youth*.

Cohen, J., McCabe, E. M., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record*, 111(1), 180-213.

Cornell, D., Shukla, K., & Konold, T. (2015). Peer victimization and authoritative school climate: A multilevel approach. *Journal of Educational Psychology, 107*, 1186-1201. doi:10.1037/edu0000038

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York, NY: Wiley.

Cronbach, L. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*, 391-418. doi:10.1177/0013164404266386

Cronbach, L. J., Linn, R. L., Brennan, R. L., & Haertel, E. H. (1997). Generalizability analysis for performance assessments of student achievement or school effectiveness. *Educational and Psychological Measurement, 57*, 373-399. doi:10.1177/0013164497057003001

Croninger, R. G., & Lee, V. E. (2001). Social capital and dropping out of high school: Benefits to at-risk students of teachers' support and guidance. *Teachers College Record, 103,* 548-581. doi:10.1111/0161-4681.00127

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*, 247-267. doi:10.1016/j.jsp.2010.02.002

Darensbourg, A. M., & Blake, J. J. (2014). Examining the academic achievement of black adolescents: Importance of peer and parental influences. *Journal of Black Psychology*, *40*, 191-212. doi:10.1177/0095798413481384

DeVellis, R. F. (2016). *Scale development: Theory and applications.* New York, NY: Sage.

Ding, C., & Hall, A. (2007). Gender, ethnicity, and grade differences in perceptions of school experiences among adolescents. *Studies in Educational Evaluation, 33*(2), 159-174.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation, 14*(20), 1-11.

Espelage, D. L., Low, S. K., & Jimerson, S. R. (2014). Understanding school climate, aggression, peer victimization, and bully perpetration: Contemporary science, practice, and policy. *School Psychology Quarterly, 29*, 233-237. doi:10.1037/spq0000090

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59-109. doi:10.3102/00346543074001059

Furlong, M. J., Greif, J. L., Bates, M. P., Whipple, A. D., Jimenez, T. C., & Morrison, R. (2005). Development of the california school climate and safety survey-short form. *Psychology in the Schools, 42*, 137-149. doi:10.1002/pits.20053

Gage, N. A., Prykanowski, D. A., & Larson, A. (2014). School climate and bullying victimization: A latent class growth model analysis. *School Psychology Quarterly, 29*, 256-271. doi:10.1037/spq0000064

Gerlinger, J., & Wo, J. C. (2016). Preventing school bullying: Should schools prioritize an authoritative school discipline approach over security measures? *Journal of School Violence, 15*, 133-157. doi:10.1080/15388220.2014.956321

Gill, M. G., Ashton, P., & Algina, J. (2004). Authoritative schools: A test of a model to resolve the school effectiveness debate. *Contemporary Educational Psychology, 29*(4), 389-409.

Gottfredson, D. C., Gottfredson, G. D., & Hybl, L. G. (1993). Managing adolescent behavior a multiyear, multischool study. *American Educational Research Journal, 30*, 179-215. doi:10.3102/00028312030001179

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

Grayson, J. L., & Alvarez, H. K. (2008). School climate factors relating to teacher burnout: A mediator model. *Teaching and Teacher Education, 24*, 1349-1363. doi:10.1016/j.tate.2007.06.005

Green, C. W., Adams, A. M., & Turner, C. W. (1988). Development and validation of the school interracial climate scale. *American Journal of Community Psychology, 16*(2), 241-259.

Gregory, A., Cornell, D., 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*, 483. doi:10.1037/a0018562

Gregory, A., Allen, J. P., Mikami, A. Y., Hafen, C. A., & Pianta, R. C. (2014). Effects of a professional development program on behavioral engagement of students in middle and high school. *Psychology in the Schools, 51*, 143-163. doi:10.1002/pits.21741

Gregory, A., & Cornell, D. (2009). "Tolerating" adolescent needs: Moving beyond zero tolerance policies in high school. *Theory into Practice, 48*(2), 106-113.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6*(4), 430.

Griffith, J. (1995). An empirical examination of a model of social climate in elementary schools. *Basic and Applied Social Psychology, 17*(1 and 2), 97-117.

Griffith, J. (2000). School climate as group evaluation and group consensus: Student and parent perceptions of the elementary school environment. *The Elementary School Journal, 101*, 35-61. doi:10.1086/499658

Haberman, S. J. (2008). When can subscores have value? *Journal of Educational and Behavioral Statistics*, 33, 204–229. doi: 10.3102/1076998607302636

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*, 177.

Hoffman, C. C., Olson, D., & Haase, S. L. (2001). Contrasting a 360-degree feedback measure with behaviorally-based assessment tools: An application of generalizability theory. *Psychologist-Manager Journal, 5*(1), 59-72.

Hoy, W. K., & Hannum, J. W. (1997). Middle school climate: An empirical assessment of organizational health and student achievement. *Educational Administration Quarterly*, 33, 290-311. doi:10.1177/0013161X97033003003

Huang, F. L. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *The Journal of Experimental Education, 84*, 175-196. doi: 10.1080/00220973.2014.952397

Huang, F. L., & Cornell, D. G. (2016). Using multilevel factor analysis with clustered data: Investigating the factor structure of the Positive Values Scale. *Journal of Psychoeducational Assessment, 34,* 3-14. doi: 10.1177/0734282915570278

Huang, F. L., Cornell, D. G., Konold, T., Meyer, J. P., Lacey, A., Nekvasil, E. K., . . . Shukla, K. D. (2015). Multilevel factor structure and concurrent validity of the teacher version of the authoritative school climate survey. *Journal of School Health, 85*(12), 843-851.

Huang, J. (2012). Using generalizability theory to examine the accuracy and validity of large-scale ESL writing assessment. *Assessing Writing, 17,* 123-139. doi:10.1016/j.asw.2011.12.003

Hughes, J. N., Cavell, T. A., & Willson, V. (2001). Further support for the developmental significance of the quality of the teacher–student relationship. *Journal of School Psychology, 39,* 289-301. doi:10.1016/S0022-4405(01)00074-7

Hurd, N. M., Hussain, S., & Bradshaw, C. P. (2018). School disorder, school connectedness, and psychosocial outcomes: moderation by a supportive figure in the school. *Youth & Society, 50,* 328-350. doi: 10.1177/0044118X15598029

Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2015). Exploring the causal impact of the McREL balanced leadership program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation & Policy Analysis, 37,* 314-332. doi:10.3102/0162373714549620

Jia, Y., Konold, T. R., & Cornell, D. (2015). Authoritative school climate and high school dropout rates. School Psychology Quarterly, doi:10.1037/spq0000139

Jiang, Z., & Raymond, M. (2018). The use of multivariate generalizability theory to evaluate the quality of subscores. *Applied psychological measurement, 42,* 595-612. doi: 10.1177/0146621618758698

Klein, J. (2012). Validity of self-report in assessing school climate and risk behavior

Konold, T., & Cornell, D. (2015). Measurement and structural relations of an authoritative school climate model: A multi-level latent variable investigation. *Journal of School Psychology, 53,* 447-461. doi:10.1016/j.jsp.2015.09.001

Konold, T., Cornell, D., Huang, F., Meyer, P., Lacey, A., Nekvasil, E., . . . Shukla, K. (2014). Multilevel multi-informant structure of the authoritative school climate survey. *School Psychology Quarterly*, *29*, 238-255. doi:10.1037/spq0000062

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.

Kottkamp, R. B., Mulhern, J. A., & Hoy, W. K. (1987). Secondary school climate: A revision of the OCDQ. *Educational Administration Quarterly, 23*, 31-48. doi:10.1177/0013161X87023003003

Lamborn, S. D., Mounts, N. S., Steinberg, L., & Dornbusch, S. M. (1991). Patterns of competence and adjustment among adolescents from authoritative, authoritarian, indulgent, and neglectful families. *Child Development, 62*(5), 1049-1065.

Lee, S., Olszewski-Kubilius, P., & Thomson, D. T. (2012). Academically gifted students' perceived interpersonal competence and peer relationships. *Gifted Child Quarterly*, *56*(2), 90-104. doi:10.1177/0016986212442568

Lewin, K. (1939). Field theory and experiment in social psychology: concepts and methods. *American Journal of Sociology, 44*, 868-896. doi: 10.1086/218177

Liu, Y., Ding, C., Berkowitz, M. W., & Bier, M. C. (2014). A psychometric evaluation of a revised school climate teacher survey. *Canadian Journal of School Psychology*, *29*, 54-67. doi:10.1177/0829573514521777

Mashburn, A. J., Downer, J. T., Rivers, S. E., Brackett, M. a., & Martinez, A. (2014). Improving the power of an efficacy study of a social and emotional learning program: Application of generalizability theory to the measurement of classroom-level outcomes. *Prevention Science, 15*, 146-155. doi:10.1007/s11121-012-0357-3

Moos, R. H. (1973). Conceptualizations of human environments. *American Psychologist*, *28*(8), 652.

Morin, H. K., Bradshaw, C. P., & Berg, J. K. (2015). Examining the link between peer victimization and adjustment problems in adolescents: *The role of connectedness and parent engagement. Psychology of Violence, 5*(4), 422.

Nelson, S. E., & Dishion, T. J. (2004). From boys to men: Predicting adult adaptation from middle childhood sociometric status. *Development and Psychopathology, 16*(02), 441-459.

Olweus, D. (1994). Annotation: Bullying at school: Basic facts and effects of a school based intervention program. *Child Psychology & Psychiatry & Allied Disciplines, 35*(7), 1171-1190. doi:10.1111/j.1469-7610.1994.tb01229.x

Pellerin, L. A. (2005). Applying Baumrind's parenting typology to high schools: Toward a middle-range theory of authoritative socialization. *Social Science Research, 34,* 283-303. doi:10.1016/j.ssresearch.2004.02.003

Perry, A. C. (1919). *The management of a city school.* New York, NY: Macmillan.

Peters, G. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist, 16*(2), 56-69.

Pianta, R. C., Stuhlman, M. W., & Hamre, B. K. (2002). How schools can do better: Fostering stronger connections between teachers and students. *New Directions for Youth Development*, 91-107. doi: 10.1002/yd.23320029307

Piscatelli, J., & Lee, C. (2011). *State policies on school climate and bully prevention efforts: Challenges and opportunities for deepening state policy support for safe and civil schools*. New York, NY: National School Climate Center,

Robinson, W. S. (2009). Ecological correlations and the behavior of individuals. *International Journal of Epidemiology, 38*, 337-341. doi:10.1093/ije/dyn357

Schweig, J. D. (2014). Quantifying error in survey measures of school and classroom environments. *Applied Measurement in Education, 27,* 133-157. doi:10.1080/08957347.2014.880442

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Thousand Oaks, CA: Sage.

Shear, L., Means, B., Mitchell, K., House, A., Gorges, T., Joshi, A., . . . Shkolnik, J. (2008). Contrasting paths to small-school reform: Results of a 5-year evaluation of the bill & Melinda Gates foundation's national high schools initiative. *Teachers College Record, 110,* 1986-2039.

Steinberg, L., Lamborn, S. D., Dornbusch, S. M., & Darling, N. (1992). Impact of parenting practices on adolescent achievement: Authoritative parenting, school involvement, and encouragement to succeed. *Child Development., 63*(5), 1266-1281.

Stockard, J., & Mayberry, M. (1992). *Effective educational environments*. Newbury Park, CA: Corwin.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*, 99-103. doi:10.1207/S15327752JPA8001_18

Sugai, G., & Horner, R. H. (2009). Responsiveness-to-intervention and school-wide positive behavior supports: Integration of multi-tiered system approaches. *Exceptionality, 17*, 223-237. doi:10.1080/09362830903235375

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics (5th ed.)*. Boston, MA: Allyn & Bacon/Pearson Education.

Taylor, L. C., Hinton, I. D., & Wilson, M. N. (1995). Parental influences on academic performance in african-american students. *Journal of Child and Family Studies, 4*, 293-302. doi:10.1007/BF02233964

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

Towles-Reeves, E., Kearns, J., Flowers, C., Hart, L., Kerbel, A., Kleinert, H., Quenemoen, R., & Thurlow, M. (2012). *Learner characteristics inventory project report (A product of the NCSC validity evaluation)*. Minneapolis, MN: University of Minnesota, National Center and State Collaborative.

Turner, I., Reynolds, K. J., Lee, E., Subasic, E., & Bromhead, D. (2014). Well-being, school climate, and the social identity process: A latent growth model study of bullying perpetration and peer victimization. *School Psychology Quarterly, 29*, 320-335. doi:10.1037/spq0000074

Vacha-Haase, T., Henson, R. K., & Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement, 62*, 562-569. doi:10.1177/0013164402062004002

Volpe, R. J., McConaughy, S. H., & Hintze, J. M. (2009). Generalizability of classroom behavior problem and on-task scores from the direct observation form. *School Psychology Review, 38*(3), 382-401.

Wang, M., & Dishion, T. J. (2012). The trajectories of adolescents' perceptions of school climate, deviant peer affiliation, and behavioral problems during the middle school years. Journal of Research on Adolescence : *The Official Journal of the Society for Research on Adolescence, 22*(1), 40-53.

Wang, M., Selman, R. L., Dishion, T. J., & Stormshak, E. A. (2010). A tobit regression analysis of the covariation between middle school students' perceived school climate and behavioral problems. *Journal of Research on Adolescence, 20*, 274-286. doi:10.1111/j.1532-7795.2010.00648.x

Way, N., Reddy, R., & Rhodes, J. (2007). Students' perceptions of school climate during the middle school years: Associations with trajectories of psychological and behavioral adjustment. *American Journal of Community Psychology, 40*, 194-213. doi:10.1007/s10464-007-9143-y

Webb, N. M., Shavelson, R. J., & Haertel, E. H. (2006). 4 reliability coefficients and generalizability theory. *Handbook of Statistics, 26*, 81-124. doi: 10.1016/S0169-7161(06)26004-8

Wentzel, K. R. (2006). *A social motivation perspective for classroom management*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Yin, Y., & Shavelson, R. J. (2008). Application of generalizability theory to concept map assessment research. *Applied Measurement in Education, 21*, 273-291.

You, S., O'Malley, M. D., & Furlong, M. J. (2014). Preliminary development of the Brief–California school climate survey: Dimensionality and measurement invariance across teachers and administrators. *School Effectiveness and School Improvement, 25,* 153-173. doi:10.1080/09243453.2013.784199

Zullig, K. J., Collins, R., Ghani, N., Patton, J. M., Scott Huebner, E., & Ajamie, J. (2014). Psychometric support of the school climate measure in a large, diverse sample of adolescents: A replication and extension. *Journal of School Health, 84,* 82-90. doi: 10.1111/josh.12124

Zullig, K. J., Collins, R., Ghani, N., Hunter, A. A., Patton, J. M., Huebner, E. S., & Zhang, J. (2015). Preliminary development of a revised version of the school climate measure. *Psychological Assessment, 27*, 1072-1081. doi:10.1037/pas0000070

Zullig, K. J., Koopman, T. M., Patton, J. M., & Ubbes, V. A. (2010). School climate: Historical review, instrument development, and school assessment. *Journal of Psychoeducational Assessment, 28,* 139-152. doi:10.1177/0734282909344205

Appendix

UNIVERSITY OF DELAWARE

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE:	February 11, 2016
TO:	George Bear, Ph.D.
FROM:	University of Delaware IRB
STUDY TITLE:	[161809-5] School Climate in Delaware Public Schools
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	DETERMINATION OF EXEMPT STATUS
DECISION DATE:	February 11, 2016
REVIEW CATEGORY:	Exemption category # (4)

Thank you for your submission of Continuing Review/Progress Report materials for this research study. The University of Delaware IRB has determined this project is EXEMPT FROM IRB REVIEW according to federal regulations.

We will put a copy of this correspondence on file in our office. Please remember to notify us if you make any substantial changes to the project.

If you have any questions, please contact Nicole Farnese-McFarlane at (302) 831-1119 or nicolefm@udel.edu. Please include your study title and reference number in all correspondence with this office.

cc: