Applications of Latent Class Models:

Profiles of School Climate and Invalid Respondents in Self-Reports

---

A Dissertation

Presented to

The Faculty of the Curry School of Education

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Kathan Shukla, M.Sc., M.Ed.

August 2015

Abstract

Advisor: Timothy R. Konold,
Professor and Program Director of Research, Statistics, and Evaluation,
Curry School of Education.

Latent class (LC) modeling is a model based cluster analysis technique. This dissertation applied LC modeling techniques to a measurement problem (study 1) and a substantive problem (study 2). It used secondary data from the Virginia Secondary School Climate Survey (VSSCS; Cornell et al., 2014). Study 1 contained 52,012 students from 323 high schools; whereas the second study sample consisted of 47,631 students from 323 high schools.

The first study introduced a novel technique for identifying invalid respondents in self-reported questionnaires (SAQs). Respondent characteristics such as joking, lying, and/or responding carelessly could undermine the validity of the study. It is desirable to screen out such invalid respondents from the analytic sample for accurate inferences. The proposed technique was conducted in three steps: 1) creation of a response-inconsistency variable that gauged the extent to which the individual's responses increased or decreased the coefficient alpha for the sample for each scale, 2) application of latent class modeling on these variables to examine the clustering at extreme values of the response-inconsistency variables, and 3) cross-validation of cases identified as invalid with traditionally used techniques like screening item (additional item asking students if they responded truthfully) and response time data (if students reported too fast so that they

may not have read the questions carefully). Researchers across different fields of social sciences may find this technique useful.

The second study adopted a comprehensive person-centered analytic approach through the examination of profiles of student perceptions of school climate in high schools. Multilevel analyses helped unpack four meaningfully different within-school (student-level) latent profiles: positive climate class, medium climate-low bullying class, medium climate-high bullying class, and negative climate class. On average, students reporting higher levels of disciplinary structure, academic press, teacher respect for students, student’s willingness to seek help from teachers, academic engagement and cognitive engagement also reported lower levels of PTB, general victimization, and probability of being bullied and bullying others. In addition, students reporting positive climate also reported higher academic outcomes and lower risk behaviors. Finally, the implications of within-school clusters of school climate are discussed.

ACKNOWLEDGEMENTS

I would not have been able to complete this dissertation without the learning and research environment that the University of Virginia provided, the guidance of my committee members, and the support from my friends and family.

My advisor, Dr. Timothy Konold, has been a great mentor and a colleague. I express my deepest gratitude for his valuable guidance and providing me with a supportive atmosphere for pursuing research. I would like to thank Dr. Dewey Cornell for his insightful feedback, and allowing me to be a member of his excellent research group and to use Virginia Secondary School Climate Survey data for this work. I also thank Dr. Patrick Meyer and Dr. Ji Hoon Ryoo for helping me develop understanding in measurement and statistical modeling.

I am grateful to all friends and well-wishers, especially my family: grandparents – Vasumati and Suresh Shukla; parents – Uma and Dushyant; elder-brother & family – Chintan, Namrata, Tanay and Darsh; and younger brother – Vedant. I am greatly indebted to my parents-in-law, Jyotsna and Pinak Chakrabartty as they continue to take a great care of my wife while I am in the US. I also acknowledge Monica and Akshay's strong emotional and moral support over the years. My gratitude toward my loving wife, Alo, is beyond words. Long-distance relationship could have been impossible, but she stood by me through the good times and bad. You all have been my greatest strength.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Finite mixture modeling is a statistical technique that helps in analyzing data consisting of multiple probability distributions. Mixture models have found applications in a wide range of fields including biology, medicine, psychiatry, engineering, physical, and social sciences (McLanchlan & Peel, 2000). One of the most important reasons for this wide popularity is that mixture modeling provides a highly flexible mathematical tool for latent class (LC) analysis (latent cluster analysis). LC modeling is defined as a technique that is used to identify the underlying subgroups (i.e. latent classes) in a population across a set of theoretically selected variables. In this way, latent classes are determined from individuals' response patterns. Recent technological advancements have enabled many applied researchers and practitioners to employ LC models.

This dissertation has three primary goals: 1) to describe and review the methodological literature on latent class modeling; 2) to apply LC modeling techniques to a measurement problem (study 1); and 3) to apply LC modeling technique to a substantive problem (study 2).

For the measurement application of LC modeling (study 1), this dissertation studied latent profiles of invalid adolescent respondents (non-serious respondents, jokers, or liars) in self-administered questionnaire data. In this novel approach, an individual-level response-inconsistency variable, which gauged the extent to which an individual's responses increased or decreased the coefficient alpha for the sample, was generated for each scale in the survey. The idea was to identify latent profiles through application of LC modeling on these response-inconsistency variables. Respondents who exhibited extreme profiles on these response-inconsistency variables (clustering at extreme values of the distribution) for all study-scales were identified as invalid respondents. The resultant latent profiles were cross-validated through further inspection of their association with screening items (if students reported telling truth) and response time (if students reported too fast so that they may not have read the questions carefully). The following research questions were studied for this measurement application of LC models:

1. What are the profiles of invalid respondents?
2. What is the association between invalid respondent profiles and respondents who finish the survey too fast and admit not responding truthfully?

The substantive application of LC modeling (study 2) examined the latent profiles of school climate reported by high school students in a multilevel framework. A multidimensional school climate measure consisted of eight distinct continuous scales (disciplinary structure, academic press, students' willingness to seek help, respect for students, affective and cognitive engagement, PTB, and general victimization), and two dichotomous items asking students if they were bullied or if they bullied others during the past year. Further, the relations between these emergent latent profiles and academic and risk behavior outcomes were explored to examine the evidence for their convergent validity. The following research questions were studied for this substantive application of LC models:

3. What are the within-school (student-level) response profiles of school climate among high schools?

4. Do the emergent latent profiles of school climate differ on demographic factors (e.g., gender, race/ethnicity, grade-level, and parental educational level at student level)?

5. What is the relation between the emergent latent profiles of school climate and academic outcomes (self-reported grades and academic expectations) and risk behaviors among high school students?

This dissertation used secondary data, but presents new analyses for both studies (measurement and substantive). The primary data came from the Virginia Secondary School Climate Survey (VSSCS), which is a part of the state's annual School Safety Audit program (Cornell et al., 2014). Note that the two sets of research questions (measurement and substantive) employed somewhat different subsamples. The first set of questions used an unscreened sample that included all respondents because the study purpose was to develop a technique which identifies invalid respondents. In contrast, the substantive set of questions was examined on data screened for invalid responders to obtain more accurate substantive inferences. Invalid respondents were identified based on the new technique proposed in study 1, a screening item (students who reported they were not answering truthfully), and response time technique (students who completed the survey too quickly). In total, the measurement study (study 1) contained 52,012 students from 323 high schools; whereas the substantive study (study 2) sample consisted of 47,631 students from 323 high schools.

Significance

The first study introduced a new method for exploring patterns of invalid respondents in self-reports. Self-administered questionnaires are one of the most widely used methods across various fields of social sciences. This new method is intended to help deal with validity threats due to respondent characteristics such as joking, lying, and/or responding carelessly. Traditionally used techniques need either multiple data sources (e.g., triangulation method), additional items planned in-advance (e.g., screening items), or online survey administration (e.g., response time) for creating flags for invalid respondents. However, this newly proposed technique can be used even when none of the other techniques are feasible. In addition, when other techniques are available, this technique could provide additional information about invalid patterns and can be used for cross-validation. Researchers across different fields of social sciences may find this method useful in screening out invalid respondents in order to obtain more accurate inferences.

The second study extended our understanding of school climate by adopting a comprehensive person-centered analytic approach through examination of response profiles of school climate measures among high school students. School climate is an important construct because it refers to malleable characteristics such disciplinary structure and support to students by adults in the school that can be targeted for interventions to drive school reform efforts and achieve desirable student outcomes. Multilevel analyses helped unpack within-school (student-level) profiles of school climate. Four meaningfully different within-school student profiles were revealed: positive climate class, medium climate-low bullying class, medium climate-high bullying class, and negative climate class. On average, students reporting higher levels of disciplinary structure, academic press, teacher respect for students, student's willingness to seek help from teachers, academic engagement and cognitive engagement also reported lower levels of PTB, general victimization, and probability of being bullied and bullying others. Conversely, students reporting lower levels of structure, academic press, support, engagement tended to report higher levels of bullying climate and personal victimization in schools. Evidence for the convergent validity of resultant profiles was obtained because students reporting positive climate also reported higher academic outcomes and lower risk behaviors. Schools tend to take up highly standardized approaches for creating and maintaining systems dealing with school discipline and student support. It is important for the adults in the schools to be mindful of these meaningful within-school clusters of perceived climate across students. It may be desirable to examine a possibility for a graded supportive environment, especially by providing higher support and clearer (and more consistent) disciplinary structure for students at risk for negative academic and behavioral outcomes.

Chapter 2: Literature Review

This chapter is organized around the following three themes:

1. Latent Class (LC) Modeling and its Applications

This section reviews literature on the development of LC models, LC modeling specifications, parameter estimation procedures, model fit and selection criteria, and potential application problems. This section concludes with the description of some of the advantages of LC models over standard cluster analysis techniques.

2. Measurement Issues with Self-Administered Questionnaires

The purpose of this review is to present a case for an application of LC modeling technique to tackle a measurement issue. More specifically, this section reviews literature on factors that may threaten the validity of self-administered questionnaires (SAQs), and potential approaches for dealing with these threats. Effects of invalid respondents (jokesters, liars, mischievous and/or careless respondents) on study results are summarized and justification for an LC application for identification of these respondents is provided.

3. Patterns of School Climate in High Schools

The purpose of this section is to present literature for a substantive application of LC modeling techniques. There is a review of some of the important work on school climate, its critical dimensions, and associated student outcomes. Also, a brief summary on patterns of school climate by LC modeling application is presented. Finally, a case for application of LC modeling is presented for examining: 1) latent profiles of school climate for high school students in a multilevel framework; and 2) relations between these profiles and academic outcomes and risk behaviors.

Latent Class Modeling and its Applications

A search for meaningful groups (or clusters) in data has been one of the elemental inquiries in both natural and social sciences. Finite mixture modeling is a statistical technique which helps in analyzing and inferring a distribution of data which consists of multiple probability distributions. The term *mixture* refers to the notion that the data are sampled from multiple populations and can be described by multiple probability distributions (McLanchlan & Peel, 2000). Accordingly, one may find several clusters of data in a mixture distribution. Each cluster may demonstrate somewhat unique distribution characteristics and have its own unique set of parameter values (e.g., means, standard deviations, and size). Mixture models have found applications in a wide range of fields including biology, medicine, psychiatry, engineering, physical, and social sciences (McLanchlan & Peel, 2000). One of the most important reasons for this wide popularity is that mixture modeling provides a highly flexible mathematical tool for understanding heterogeneity that may exist within a population with respect to a certain set of variables. It should be noted that the term *mixture modeling* is more prevalent in the fields of mathematics, statistics and natural sciences. However, in educational and psychological fields, researchers are often specifically interested in finding out latent clusters of participants across variables of interest. Therefore, a more widely used term in these fields is latent class modeling.

A latent class (LC) model is defined as “a probabilistic model that represents unobserved sub-populations within an overall population based on the responses to multiple observed variables” (The Pennsylvania State University, 2010). LC modeling is defined as a technique used to identify the underlying subgroups (i.e. latent classes) in a population. Latent classes are determined from individuals' response patterns. There is a great amount of interest in this statistical method and it is widely used by social, behavioral, and health researchers. Technological advancements in computational power and programming have enabled many applied researchers and practitioners to employ LCA. Software programs like M*plus* (Muthen & Muthen, 1998-2014), R (LCA package by Ed Curry), and PROC LCA (Lanza, Dziak, Huang, Wagner, & Collins, 2013) have computerized highly complex and challenging estimation procedures. Users can specify the number of hypothesized latent classes in a model and examine their fit to a dataset. When run correctly, these software packages can produce model fit statistics, latent class membership probabilities for the each sample-case, and class-specific estimates of model parameters.

As interest in LC modeling continues to grow, this section attempts to review methodological literature in a non-mathematical language. First it describes how LC models are connected with other latent variable models. Then there is a brief historical overview, and discussion of LC model specifications (statistical assumptions, parameter estimation, and model fit and information criteria) and the challenges that researchers usually face while applying LC modeling. Finally, this section describes how LC modeling compares to other traditional cluster analysis techniques.

Family of Latent Variable Models

In statistical analysis, latent variable models provide a vital mechanism for dealing with multivariate data. Latent variables can be continuous (e.g., factor analysis, and latent trait analysis) or categorical (e.g., latent class analysis, and latent profile analysis). Bartholomew, Knott, and Moustaki (2011) discuss two major reasons for widespread applications of continuous latent variables. First, latent variables help reduce dimensionality in multivariate space. The idea is to extract common variances across all indicator variables and represent it by a latent variable (i.e., factor) without significant loss of information. For example, a questionnaire on school safety may have 500 respondents and 70 items. If the researcher only reports the descriptive results, it may be difficult for the readers to meaningfully interpret information for 70 variables. If the same data can be presented with four or five meaningfully different latent variables, the readers may be able to see the overarching patterns. Therefore, researchers often strive for reducing dimensionality and conveying their work in a parsimonious manner through application of latent variables. The second major reason deals with measurability of the constructs of study interest. Unlike natural sciences, most of the study constructs are not directly measurable in social sciences, especially in education (e.g., school climate, quality of teaching, or student learning) and psychology (e.g., motivation, intelligence, or depression). Latent variable models provide a mechanism to study these unmeasurable constructs, which serve as latent variables, through indirect measurable indicator variables. For example, student reports of disciplinary structure and support from teachers in their school may serve as indicators for measuring a latent construct of school climate.

Categorical latent variables are primarily considered for identifying subgroups of objects (or study participants) in data. More about their applications is discussed in the later part of this section. In total, latent variables help identify trait-centered or person-centered patterns (or groups) depending upon which model is employed.

It is important to note that in the methodological literature a statistical model may have different names depending upon a researcher's terminology and area of research. In general, a statistical model that presents relations between a latent variable and its indicators is termed a measurement model. When a model consists of a relation(s) between two or more continuous latent variables, it is called a structural model (Klein, 2011). Models consisting of categorical latent variables are called latent class models or mixture models. Finally, a statistical model that integrates continuous and categorical latent variables is generally referred to as a structural equation mixture model.

Bartholomew et al. (2011) presented a unified approach to latent variable models that consists of factor analysis, latent trait analysis, and latent class models. Table 1 presents the classification of latent variable models.

Table 1

Classification of Latent Variable Models

		Observed Variables	
		Continuous	Categorical
Latent Variables	Continuous	Factor Analysis	Latent Trait Analysis (or Item Response Theory)
	Categorical	Latent Profile Analysis	Latent Class Analysis

When continuous indicators are represented by a continuous latent variable, it is called factor analysis. When categorical indicator variables are treated as a continuous variable, it is referred to as latent trait analysis (or item response theory). When a categorical latent variable has categorical indicators, it is termed a latent class analysis (LCA); and when its indicators are continuous, the modeling is called latent profile analysis (LPA). However, a distinction between LCA and LPA is typically not necessary given the fact that they are essentially the same models with the same purpose of examining subgroups of participants (Pastor, Barron, Miller, and Davis, 2007). Furthermore, modern software packages (e.g., M*plus*) allow both categorical and continuous indicators to be analyzed simultaneously. Accordingly, this dissertation used term Latent Class modeling which refers to both LCA and LPA. Note that in the substantive application (study 2), both continuous and categorical indicators were used as indicators for latent classes.

Conceptually, the idea of using continuous multivariate outcomes for classifying study participants into groups resembles discriminant function analysis (DFA). However, both approaches differ fundamentally because the groups are known in DFA, but unknown in LC models.

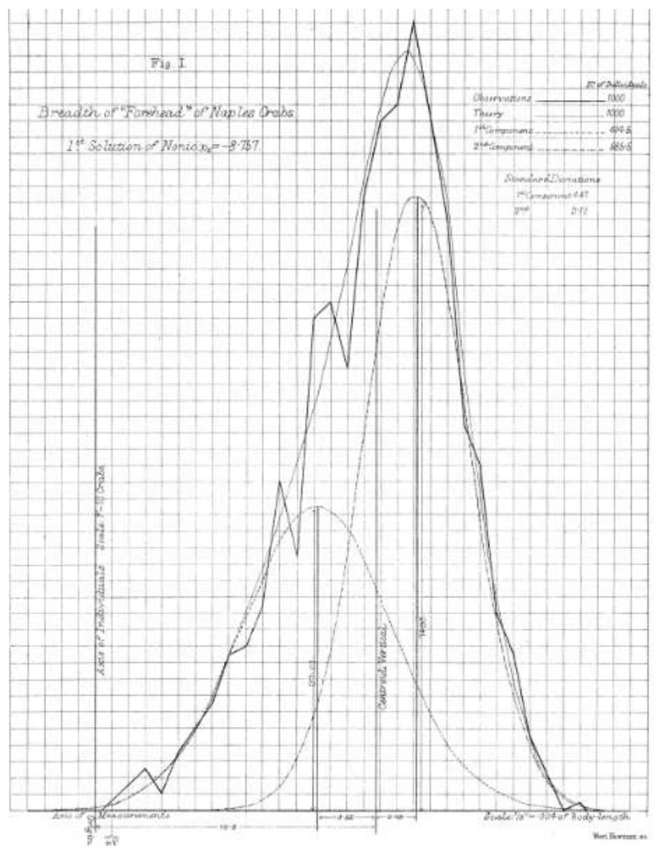

Figure 1 Breadth of "Forehead" of Naples Crabs. Source: Pearson (1894)

Brief History of LC Models

Over a century ago, Karl Pearson's (1896) manuscript entitled, "Contributions to the Mathematical Theory of Evolution", introduced a procedure for estimating a mixture model. In this paper, he employed a method of moments approach to the estimation of the five parameters (two means, two variances, and mixing proportion) in a two component univariate Gaussian mixture distribution of the measurement of Naples crab foreheads. It was a laborious act that required solving 9^{th} degree polynomial equations with many unknown quantities by hand.

Given the computational challenges associated with parameter estimations, the major methodological developments in mixture modeling were slow despite its potential for application in many fields.

In a fine discussion of the history of mixture models, Everitt (1996) wrote:

> During the next 30 years [after Pearson's work] there were a number of other attempts to simplify Pearson's proposed method. These included the use of cumulants rather than moments by Stromgren and the use of k-statistics by Rao. Despite the computational problems, associated with the two-component, univariate Gaussian mixture, Charlier and Wicksell attempted the estimation of the parameters in a two component mixture of bivariate normal, and Doetsch considered the problem of normal mixtures with more than two components. In each case the method of moments was the estimation procedure used (p. 109).

One of the biggest technical breakthroughs came in the latter part of the 1940s when Radhakrishna Rao (1948) suggested the maximum likelihood (ML) estimation for the normal mixture problem. Rao developed an iterative solution for the case of two mixture components, with equal standard deviations. Hasselblad (1966, 1969) explored the application of ML for more general mixture distributions from the exponential family (e.g., Poisson, binomial and exponential distributions). Hasselblad's work also allowed for more than two distributions and for components to have unequal variances. Cohen (1967) examined the conditional maximum likelihood estimates with first four sample moments equated to corresponding population moments. Wolfe (1967, 1970) employed maximum likelihood estimation to the situation involving multivariate normal density mixtures. Furthermore, he developed the first computer program that made the application of such mixture models possible on a routine basis. The procedure suggested by Wolfe prepared the ground for Dempster, Laird, and Rubin (1977), who later formulated the EM (Expectation Maximization) algorithm in more general terms.

It is important to note that the term "latent class analysis" was introduced by Lazarsfeld and Henry (1968), who used the LC technique to identify clusters from dichotomous indicator variables. Later, Goodman (1974) extended LC modeling to polytomous indicator variables and developed an algorithm for finding model parameter estimates from the ML method. Hagenaars explained latent class models that included local dependencies between the indicators in 1988 and latent class models from a log-linear point of view in 1993 (as cited in Collins and Lanza, 2010).

The methodological literature is continuously developing because LCA remains an area of research interest. Using National Longitudinal Survey of Youth (NLSY) data, Muthen and Muthen (2000) used growth mixture modeling to examine person-centered developmental trajectories and latent class membership across the years. Vermunt (2003, 2008) developed Multilevel LCA (MLCA) models in order to take into account nested data structures. Drawing data from 10,772 females living in 1 of 206 rural communities across the United States, Henry and Muthen (2010) demonstrated the application of LCA in a multilevel setting with individuals at level-1 and communities at level-2. Their best-fitting model consisted of three level-1 latent classes (heavy smokers, moderate smokers, and nonsmokers), two random effects to capture the uncertainty regarding level-1 class membership, and a random factor for level 2 indicators. In order to distinguish between unobserved heterogeneity at level-1 and level-2 heterogeneity in the form of random effects, Muthen and Asparouhov (2009) developed the multilevel regression mixture analysis. In other words, they developed the methodology for testing hierarchical linear models in different unobserved subpopulations. Researchers often describe LCA models as a categorical analogy to factor analysis, where the latent variable is continuous. Lubke and Muthen (2005) combined both modeling approaches (LCA and factor analysis) to develop factor mixture models (FMM). FMMs provide a tool for examining population heterogeneity and can describe unobserved clusters of study participants within a factor analysis framework.

The Latent Class Model

The LC model is a mixture model consisting of two or more probability distributions. The most basic model consists of dichotomous indicator variables. Vermunt and Magidson (2004) express this mathematically as follows:

$$P(\mathrm{Y=y}) = \sum_{x=1}^{C} P(\mathrm{X} = x)\, P(\mathrm{Y} = \mathrm{y} \mid \mathrm{X} = x) \qquad (1)$$

Where, y= response pattern; $P(\mathrm{Y=y})$ =probability of obtaining response pattern,

$P(\mathrm{X}=x)$ = proportion of persons in latent class ‘x’

Equation (1) denotes that the probability of getting a response pattern y, is a weighted average of the x class-specific probabilities, $P(\mathrm{Y=y} \mid \mathrm{X}=x)$.

Note that the latent variable may have multiple indicator variables (L). Therefore, class-specific probabilities of a response pattern, $P(\mathrm{Y=y} \mid \mathrm{X}=x)$, can be obtained as the product of the probability of obtaining a response pattern y for an indicator (l) for a particular class (x) across all indicators.

$$P(\mathrm{Y=y} \mid \mathrm{X}=x) = \prod_{l=1}^{L} P(Y_l = y_l \mid \mathrm{X} = x) \quad (2)$$

Here, indicators (L) are assumed to be mutually independent within classes. More on the local independence assumption will be presented in the sub-section to follow.

By substituting the value of $P(\mathrm{Y=y} \mid \mathrm{X}=x)$ from equation 2 to equation 1, one gets –

$$P(\mathrm{Y=y}) = \sum_{x=1}^{C} P(\mathrm{X} = x) \prod_{l=1}^{L} P(Y_l = y_l \mid \mathrm{X} = x) \quad (3)$$

One of the prime objectives of LCA is to attain the best possible solution for the above expression.

The Latent Profile Analysis (LPA; LCA with continuous indicators) is an extension of the basic LCA. In LPA, the response probabilities are replaced with response densities (Vermunt 2004; Vermunt & Magidson, 2004). Therefore, equation 2 is denoted by the form of:

$$f(\mathrm{y}) = \sum_{x=1}^{C} P(x) f(\mathrm{y} \mid \mu_x, \Sigma_x) \quad (4)$$

Equation (4) indicates that the joint density of L indicators, $f(\mathrm{y})$, can be represented by a mixture of class(x)-specific densities. Also, μ_x and Σ_x are the mean and covariance matrix of class x. $P(x)$ is the proportion of persons in class x.

LC modeling is quite flexible in a sense that the class mean and covariance matrix can be constrained across classes in a variety of ways. When the number of free parameters (parameters to be estimated) are more than the number of unknown unique elements in the covariance matrix, the statistical model is said to be unidentified and it fails to converge to the global maximum of the log-likelihood function. Accordingly, the model fails to generate accurate parameter estimates. Model restrictions may help deal with such model-convergence problems and conduct analysis in a parsimonious manner because there are limited number of unknown parameters to estimate. Vermunt (2004) explained –"Several special cases are obtained by restricting the covariance matrix Σ*x*. Common restrictions are equal covariance matrices across classes, diagonal covariance matrices, and both equal and diagonal covariance matrices".

When Local Independence across indicators is assumed, one gets diagonal covariance matrices. Accordingly, equation (3) can be expressed as below for the continuous indicators:

$$f(\mathrm{y}) = \sum_{x=1}^{C} P(x) \prod_{l=1}^{L} f(y_l|\mu_{lx}, \sigma^2{}_{lx}) \tag{5}$$

Where, $\sigma^2{}_{lx}$ represents variance of indicator *l* in class *x*. All covariances in the covariance matrix Σ_x are constrained to be zero.

Pastor et al. (2007) presented five approaches for imposing successively more severe? restrictions on covariance matrices: 1) all variances and covariances are allowed to vary across indicators and across classes; 2) covariances are allowed to vary across indicators within classes, but restricted across classes; 3) all covariances are constrained to zero; 4) all variances and covariances are allowed to vary across indicators within classes (and restricted across classes); and 5) variances are allowed to vary across indicators within classes while all covariances are constrained to zero (i.e., indicators assumed to be unrelated). As the researcher moves from model 1 to 5, the number of free parameters associated with covariance matrix decreases.

Statistical Assumptions

Collins and Lanza (2010) have written a seminal introductory book on latent class analysis and provide a nice explanation of statistical assumptions. In LC modeling, the level of measurement of the observed variables or the latent classes is not assumed to be continuous. Therefore, the indicator variables can be categorical and their joint distribution can be multinomial. Accordingly, multivariate normality is not assumed.

Local independence. The local independence assumption states that the observed variables are independent of one another after controlling for latent variable. In other words, within each latent class, the observed variables would be independent. For example, Collins and Lanza (2010) presented the following Figure 2-A and 2-B.

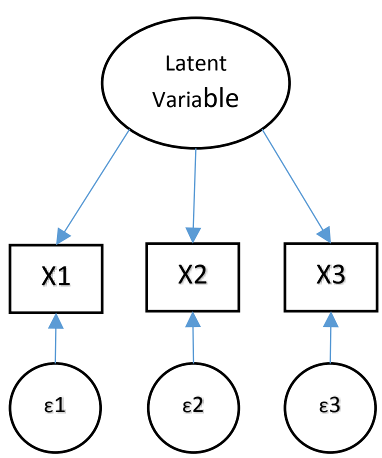

Figure 2-A

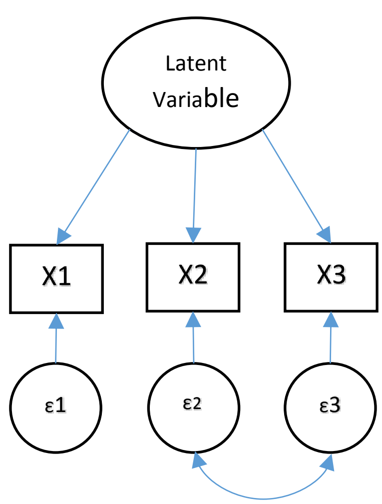

Figure 2-B

Figure 2-A and 2-B indicate a latent variable with three indicator variables X1, X2, and X3. ε1, ε2, and ε3 are error terms associated with X1, X2, and X3, respectively. In Figure 2-A, errors are not related with one another. It indicates that the observed variables are interrelated only through the latent variable. However, in Figure 2-B, there is a double arrow linking error terms ε2 and ε3. This indicates that X2 and X3 are related with each other even after controlling for the latent variable. Thus, Figure 2-B represents an example of a violation of the local independence assumption.

It is important to note that in a dataset the observed variables can be interrelated. Observed data can be a mixture of multiple latent classes and local independence refers to the independence of variables within these classes.

Parameter estimation

There are primarily two estimation approaches for latent class analysis: the maximum likelihood (ML) approach, and the Bayesian approach. With ML, the objective is to maximize the likelihood for unknown parameter values for a known population distribution for given data. ML estimation typically uses an iterative procedure like the Newton-Raphson (Agresti, 2002) or EM (Expectation Maximization; Dempster et al., 1977) algorithms. Agresti (2002) explains the Newton-Raphson procedure in a lucid manner as follows:

> It begins with an initial guess for the solution. It obtains a second guess by approximating the function to be maximized in a neighborhood of the initial guess by a second-degree polynomial and then finding the location of that polynomial's maximum value. It then approximates the function in a neighborhood of the

> second guess by another second-degree polynomial, and the third guess is the location of its maximum. In this manner, the method generates a sequence of guesses. These converge to the location of the maximum when the function is suitable and/or the initial guess is good (p.g., 143-144).

The EM algorithm iterates between the expectation step and the maximization step. Iteration continues until the convergence criterion is met. Like the Newton-Raphson procedure, the convergence criterion is often some small difference in the value of parameter estimates between successive iterations. Note that EM algorithm is often a quicker way of obtaining a solution than the Newton-Raphson procedure. However, the EM algorithm is more likely to skip global maximum while it is in the expectation step and converge at a local maximum of the data likelihood function in a multimodal distribution if the starting value is inappropriate.

The Bayesian approach relies on Bayes' Theorem and is used to find the joint probability distribution of parameters for given data. Note that analytically it can be difficult to find the posterior distribution of parameters. Researchers have suggested the application of Markov chain Monte Carlo (MCMC) methods, especially Gibbs sampling, to achieve convergence to the posterior distribution (for details, Diebolt, & Robert, 1994).

Model Fit and Model Selection

As discussed in the beginning, the number of classes are not known in most applications of LCA. Therefore, in order to identify the best fitting model that explains the data and describes the heterogeneity optimally, it is important to run a series of LCA models with different classes and compare their model fit statistics. In most cases, this exploration for the best fitting model starts with a 2-class LCA model. Thereafter, the number of classes are gradually increased by one at a time until the model fails to converge or the results no longer make sense. Finally, the model fit statistics are compared to determine the number of classes and identify a model that provides the most meaningful and statistically valid results. Model fit statistics are examined mainly through information criteria.

Information criteria. Numerous indices have been suggested by researchers over the years. It is often recommended to report various fit indices as there is no consensus in the literature over any single model fit index. In the context of LCA, the most widely reported fit indices are AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), LRT (Likelihood-Ratio Test), VLMR LRT (Vuong, Lo, Mendell, Rubin – LRT), LMR LRT (Lo, Mendell, & Rubin – LRT), BLRT (Bootstrapping LRT). Lower values of AIC and BIC indicate a better model fit.

AIC. AIC is often used in mixture modeling. However, it is often found to over-extract the number of classes (Nylund, Asparouhov, & Muthen 2007). It is defined as below:

$$AIC = -2 \ln L + 2p$$

Where, p= number of parameters; & L = likelihood function.

-2 ln L is called deviance, which is a measure of misfit.

There is no specific range of values for AIC and no prescribed cut-off value that reflects a good model fit. Nonetheless, lower AIC values are considered a good source for evaluating the relative fit of competing models.

BIC. Some simulation studies have found BIC to be the best of the information criteria for extracting the correct number of classes (e.g., Nylund, Asparouhov, & Muthén, 2007).

$$BIC = -2 \ln L + p \ln(N)$$

Where, N= sample size; p= number of parameters; & L = likelihood function.

Like the AIC, lower value of BIC are taken as evidence of a better fitting model.

Likelihood Ratio Test (LRT). LRTs are statistical tests of differences between nested models. In mixture modeling there are two different kinds of LRT applications: 1) LRT used to determine the number of classes that tests models with k versus k-1 classes; and 2) LRT performed when testing nested models within the same enumerated number of classes.

In LRT for determining the nested number of classes, the

Null hypothesis: Ho: number of classes is k-1

Alternate hypothesis: H1: number of latent classes is k

$$\begin{aligned} LRT &= -2[\ln L(\text{model } 1) - \ln L(\text{model } 2)] \\ &= -2 \ln L(\text{model } 1) + \ln L(\text{model } 2) \\ &= \text{Deviance Ho} - \text{Deviance H1} \end{aligned}$$

Where model 2 is nested within model 1.

Significant *p*-value (<0.05) suggests that the model with k classes fits the data better than the model with k-1 classes.

It should be noted that likelihood-ratio statistics for model comparison do not work well when indicator variables of latent class have sparseness. Sparseness refers to the extent to which the average expected cell count is small; and is defined by N/W, where N= sample size and W= size of the contingency table.

VLMR – LRT. Vuong (1989) proposed the VLMR (Vuong, Lo, Mendell, Rubin) – LRT which compared the data to their theoretical distribution. This test uses a weighted χ^2 distribution when the models are nested and a normal distribution when the models are not nested. A small p-value would favor the model with k classes instead of the model with k-1 classes.

LMR – LRT. Lo, Mendell, and Rubin (2001) proposed an adjustment to the VLMR LRT for the difference in the number of free parameters and sample size:

$$LMR = VLMR/ [1 + \{(P_k – P_{k-1}) \ln (N)\}^{-1}]$$

Small *p*-values indicate that the k-class model fits better to data than the k-1 class model.

BLRT. This method uses the parametric bootstrapped ratio test for determining number of classes (McLachlan, 1987). In the k class run, this test estimates both the k class and k-1 class model. The test generates about 100 samples using parameter estimates from the k-1 class model. For each generated sample, LRT statistics are estimated for both k and k-1 models. Finally, we get p-value for the LRT by comparing its value to the empirical distribution based on the bootstrap sample. A significant *p*-value indicates that k-class model is a better fit to the data than the k-1 class model.

Nylund et al. (2007) conducted an important simulation study to examine the performance of information criteria (ICs; e.g., AIC, BIC) and likelihood-based tests (e.g., LRT, LMR, BLRT) for determining the number of classes in a mixture modeling framework. Model fit statistics were examined for three different sample sizes (*n*= 200, 500, and 1000) and for three types of mixture models: latent class models, factor mixture models, and growth mixture models (GMM). They found that BIC and BLRT are the best performing IC and likelihood-based test, respectively, in determining the number of classes across various models.

Quality of Classification. Entropy is a measure of uncertainly in the classification of subjects (*N*) into latent classes (K). Here, classification tables are based on posterior class probabilities. φ_{ik} where rows are cases that have the highest probability in this class; entries φ_{ik} are averaged over cases.

Entropy is defined as,

$$E_k = 1 - \frac{\sum_{i=1}^{N}\sum_{k=1}^{K}(-\emptyset ik \ ln\emptyset)}{N \ln K}$$

Where, N= sample size, and k= class

A value close to 1 indicates excellent classification in a class and that many cases have φ_{ik} values close to 0 or 1.

Posterior Class Probabilities

After identifying the best fitting model, posterior class probabilities can be estimated. Posterior class probabilities estimate the likelihood of individuals being classified to each of the latent classes, given the individuals' response pattern on observed indicators (as suggested in Nylund, 2007). Note that the classification probability of individuals depends upon model parameters. In other words, an individual may be classified in different classes in different models. Each participant can then be classified in a class with highest probability. For example, suppose an individual has posterior probabilities of 0.80 and 0.20 for classes 1 and 2, respectively. This individual would be assigned to Class 1.

Convergent Validity of the Classes

Once the final latent class model is determined, it is often of interest to examine relations between latent classes and other external variables that were not part of the latent class model. The external variables should have theoretical associations with latent class variables in the literature. Therefore, this examination serves as convergent validity evidence for the latent class (cluster) solution. Pastor et al. (2007) described two common analytic approaches that researchers can undertake. Researchers may employ analysis of variance (ANOVA) with external outcomes and class membership as predictors. Researchers may also use posterior probabilities of class membership as predictors in multiple regression to incorporate classification accuracy into the analysis. Secondly, researchers may integrate convergent validity analysis with LC modeling along with relevant covariates in a structural equation modeling framework as demonstrated by Muthen and Muthen (1998-2014). The auxiliary (e) function in M*plus* (Muthen & Muthen, 1998-2014) tests for the quality of means across latent classes using a Wald chi-square based on posterior probability-based multiple imputations. In addition, more recent developments allow examination of relations between latent classes, covariates, and/or proximal and distal outcomes within a comprehensive modelling framework (for details, see Asparouhov & Muthen, 2014).

Dealing with Missing data

Almost all empirical data are prone to have missing values in a dataset. Enders (2010) presents a clear explanation of a classification system for missing data problems originally suggested by Rubin and colleagues. The three most commonly discussed missing data mechanisms are missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR).

When the probability of missing data in the outcome is unrelated to any other variables and the values of the outcome, the data can be considered MCAR. Data are considered MAR when the probability of missing data on a variable X is related to some other variable(s) in the analytic model, but not to the same variable X. In other words, there may be systematic missingness in the predictor variables but not in the outcome. Lastly, when the probability of missing data on the outcome is related with the values of the outcome variable itself after controlling for other variables, data are considered to be MNAR.

In order to deal with MCAR and MAR data in statistical modeling, researchers typically use either full-information maximum likelihood (FIML) or multiple imputation (MI). Discussion on both of these techniques is beyond the scope of this literature review (for details, see Enders, 2010). However, for MNAR data, both FIML and MI are prone to produce biased results. Moreover, it is not possible to empirically verify if the values on the outcome variable are MNAR or MAR.

Challenging issues in Latent Class Analysis

While working with the empirical data, researchers may often find difficulties in estimating latent class models. Some of the major issues are discussed below:

Model convergence. Researchers often face model convergence issues while running complex statistical analysis like LCA. In addition to the under-identification of the model, other possible reasons for convergence problems may include, but are not limited to, poor starting values of parameters, a mis-specified model, and/or model variables being measured on different scales. Recent work suggests that estimation and interpretation of unidentified latent class models may have high rates of misclassification and poor predictive power for estimates (Abar, & Loken, 2012) and researchers need to make sure that the number of classes is not equal to or greater than the number of indicators to avoid these under-identified modeling issues.

Local versus global solution. As previously discussed, the maximum likelihood procedure is widely used for parameter estimation in latent class models. The parameter values obtained through ML estimation are associated with the highest log-likelihood value, which is known as the global maximum. However, in reality it is fairly common for a log-likelihood surface to have multiple local maxima or more than one maximum value. In such cases, it can be difficult for the model to converge at a global maximum and to arrive at accurate parameter solutions. Therefore, it is of utmost importance in LCA to estimate the model several times with different starting values of parameters.

When the model repeatedly arrives at the same parameter estimates, it can be concluded that it converges at the global maximum.

Spurious classes. Relying solely on statistical analysis may not be the best approach for determining latent classes. The number of classes in sample must also be theoretically meaningful. Bauer and Curran (2004) identified three conditions that may lead to retention of spurious latent classes: 1) model misspecification, 2) nonnormal continuous variables, and 3) nonlinear relations among model variables. Recently, Asparouhov and Muthen (2014) described how continuous skewed distributions can be integrated in the analysis. This approach may help prevent the formation of spurious classes due to nonnormal indicator variables in LC modeling.

Application of Latent Class Cluster Analysis

Cluster analysis is “the classification of similar objects into groups, where the number of groups, as well as their forms are unknown” (Kaufman and Rousseeuw, 1990). LC modeling has increasingly been applied for cluster analysis.

Vermunt and Magidson (2002) and Pastor et al. (2007) present a fine comparative analysis between LC models and other standard cluster analysis techniques (for example, k-means, hierarchical clustering, density models, graph-based models, and fuzzy clustering techniques). Firstly, LC is a model-based clustering approach where a statistical model is hypothesized for the population from which the study sample is drawn. Therefore, the number of clusters often have clearer theoretical rationale in LC than in other standard cluster analysis methods, where the approach is more data-driven. Secondly, in LC models, parameters are estimated by the maximum likelihood method, where solutions are achieved through maximization of a log-likelihood function. Non-hierarchical cluster techniques often use a similar approach where the allocation of objects to clusters are based on some criterion. Vermunt and Magidson (2002) state – "these criteria [in non-hierarchical cluster techniques] typically involve minimizing the within-cluster variation and/or maximizing the between-cluster variation. An advantage of using a statistical model is, however, that the choice of the cluster criterion is less arbitrary. Nevertheless, the log-likelihood functions corresponding to LC cluster models may be similar to the criteria used by certain non-hierarchical cluster techniques like k-means" (*pg*. 2). Thirdly, LC modeling is more flexible than other clustering techniques because, unlike other clustering techniques, the indicators on different measurement scales do not need to be transformed to the same scale for analytic purposes. In addition, LC models accommodate indicator variables of different distributional forms within clusters; and parameter constraints can be imposed to obtain a more parsimonious model.

Given that LC modeling is a probabilistic clustering approach, it produces probability values for each object's (or participant's) membership to each cluster (or class). Conceptually, this approach is similar to that of the fuzzy clustering technique. However, Vermunt and Magidson (2002) point out– "in fuzzy clustering an object's grades of membership are the 'parameters' to be estimated (Kaufman and Rousseeuw, 1990) while in LC clustering an individual's posterior class-membership probabilities are computed from the estimated model parameters and his observed scores". Thus, it may not be possible with standard fuzzy cluster techniques to classify all of the objects (or participants) to the population from which the sample is taken, but it can be done in LC analysis.

Finally, the criteria for determining the number of clusters are more well developed in the case of LC modeling (e.g., Bayes Information Criteria, Lo, Mendell, and Rubin log-likelihood ratio test, Bootstrapping log-likelihood ratio test) than in standard clustering techniques, which still rely heavily of subjective decisions. In sum, given these advantages, an increasing number of researchers are employing LC modeling for cluster analysis.

In the past couple of decades, LC modeling has found wide application across various fields, including marketing research (Wedel & DeSarbo, 1995), medical research (Everitt, 1996; Rindskopf and Rindskopf, 1986), psychiatry (Kendler, Eaves, Walters, Neale, Heath, & Kessler, 1996), education (Aitkin, Anderson, & Hinde, 1981; Ashley, Sharkey, & Parker, 2013; Dayton 1991; Lovegrove & Cornell, 2013), and psychology (Muthen & Muthen, 2000; Nylund, Bellmore, Nishina, and Graham, 2007). Aitkin and colleagues (1981) applied LCA to examine different teaching styles practiced in the UK. Their sample consisted of 1258 primary (elementary) school teachers. The study identified 12 clusters of teachers ordered from extremely formal to extremely informal teaching styles. Three overall teaching styles (i.e., formal, mixed, and informal) were considered for analysis of covariance on student outcomes of reading, mathematics, and English. Their results suggested significant differences among styles on all three outcomes after controlling for pre-test characteristics.

Conclusion

This section reviewed literature on the development of LC models, associated modeling specifications, parameter estimation procedures, model fit and selection criteria, and potential troubling issues in its application. The advantages of LC models over standard cluster analysis techniques was also considered. The next two sections discuss measurement and substantive applications of LC modeling.

Methodological Issues with Self-Administered Questionnaires

Self-administered questionnaires (SAQ) are one of the most widely used methods for collecting data in social sciences. Unlike telephone or face-to-face interviews, SAQs are inexpensive and are unlikely to be contaminated by interviewer bias. In addition, an SAQ is more likely to provide anonymity and privacy and encourage honest responses as compared to an interview (Babbie, 2013). They have to the potential to capture a broader picture of the students' personal experiences and are especially useful for studying sensitive issues (e.g., health-risk behavior; Brener, Billy, & Grady, 2003). However, researchers across various fields of social sciences have raised a few validity concerns.

In their meta-analysis consisting of 37 independent samples with 158 effect sizes across 60,926 students, Kuncel, Crede, and Thomas (2005) examined the validity of student reported GPA. About 82.4% of high school and 54.3% of college students accurately reported their GPA. 34.5% of college and 12.3% of high school students over-reported their GPA. Kuncel et al. (2005) also found that actual academic achievement and cognitive ability strongly moderated, whereas being non-White student weakly moderated self-reported grade validity. Winne and Jamieson-Noel (2002) reported that undergraduate students at a Canadian university were positively biased about their achievement and use of study tactics. In a longitudinal study on early sexual experiences of adolescents, Lauritsen and Swicegood (1997) estimated that 32% of adolescent respondents were inconsistent over a 7-year period in reporting when their first sexual experience occurred.

Research on alcohol and substance abuse support validity claims of SAQs. Brown, Kranzler, and Boca (1992) investigated three factors that may have influenced the validity of SAQs: alcohol status at the time of interview; cognitive functioning; and the self-report data collection method. Data from 234 participants were in the form of personal interviews, SAQs, and toxicological analyses of blood and urine samples. Comparison of SAQ and interview data showed that SAQs were in 90% agreement for alcohol, 93% for cocaine, and 81% for marijuana. Moreover, cognitive function level was not associated with the validity of self-reports.

In total, despite advantages, SAQ usage requires adequate caution. It is important to recognize various threats posed to the validity of self-report and to employ appropriate methodological steps to mitigate their effects. The next sub-section reviews the literature on factors that may threaten the validity of SAQs. These factors are covered in two broad categories, technical issues and respondent characteristics.

Validity Threats: Technical Issues in the Use of SAQs

Response validity tends to be affected by a large number of instrument design factors: complexity and duration of task, clarity of instruction, item wording, question sequencing, response format, and logical flow of the instrument (Del Boca & Darkes, 2003). However, researchers have been most concerned with technical issues related to common method variance and response-shift in self-reports as discussed below.

Common method variance bias. The problem of method bias has been an area of research interest for many decades. Common method variance, which is “attributable to the measurement method rather than to the constructs the measures represent”

(Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), is considered an issue of concern in behavioral science. Method variance is seen as a problem because it is often one of the main sources of measurement error. Cote and Buckley (1987) conducted a meta-analysis of 70 Multi-Trait Multi-Method (MTMM) studies and reported that about 26.3% of method variance could be present in a typical study measure. When both independent and dependent variables are measured by the same method, the variance due to common method may be confounded with the variance shared by these variables. Thus, inferences on the relations between study variables could be biased (Campbell & Fiske, 1959). For example, a researcher examines relationships between variables X and Y through SAQs. In this case, all the method error components associated with the measurement of X (e.g., social desirability, physical/mental condition, fatigue etc.) will also be present in the measurement of Y. Therefore, measurement error of X will be related with that of Y. Thus, any conclusion regarding the association between X and Y will be misleading because their correlation will be inflated.

In the social sciences, use of common methods for all study variables is prevalent. Sackett and Larson (1990) reviewed literature in psychology and related fields and found that 51% of the studies used SAQ as either the primary or sole source of data. In addition, 39% of all studies used a questionnaire or interview methodology wherein all of the data were collected in the same measurement context. To tackle the common method variance problem, Podsakoff et al. (2003) provide several procedural remedies: obtain measures of the predictor and criterion variables from different sources, temporal, proximal, psychological, or methodological separation of measurement, protecting respondent anonymity and reducing evaluation apprehension, counterbalancing question order of the measurement of predictor and criterion variables, and improving scale items. The common method variance problem can sometimes be handled by statistical modeling techniques (for example, extracting common variance across all measures of a survey; for more details see Podsakoff et al., 2003; Podsakoff & Organ, 1986).

It is important to note that despite the inherent threat of common method variance, SAQ data may provide valid and reliable results. In his book chapter entitled, “So why ask me? Are self-report data really that bad?”, Chan (2009) presented compelling arguments in defense of self-reports. He states that there are four widely prevalent myths among researchers: validity of SAQ is a flawed idea; SAQ data cannot provide accurate parameter estimates of relations between constructs; social desirability is deeply confounded with SAQ data, and non-SAQ data are always better. Chan (2009) stated – “there is no strong evidence to lead us to conclude that self-report data are inherently flawed or that their use will always impede our ability to meaningfully interpret correlations or other parameter estimates obtained from the data” (p. 333).

In order to deal with common method variance bias, the Multi-Trait Multi-Method (MTMM) approach provides a useful mechanism (e.g., Konold & Shukla, 2013). In their examination of college students with attention-deficit/hyperactivity disorder, Konold and Glutting (2008) found that the ratings from parents and students uniquely explained method variances after controlling for trait factors. In addition, parent ratings were better measures of internalizing behavioral dimensions and student ratings were better measures of externalizing dimensions of behavior. Konold and Shukla (2014) examined discrepancies in behavioral ratings from mothers and teachers and their associations with achievement in a longitudinal setting. Their findings suggested that although behavioral ratings from both informants explained equivalent levels of reading achievement variance, teacher ratings of behavior explained more variation than mother ratings for achievement in applied problems.

In total, multiple informants may provide a more complete picture of study parameters. However, multi-informant studies are perceived to be expensive, time-consuming, and ineffective (if informants do not cooperate). Contrary to this widely prevalent belief, Vazire (2005) demonstrated that informant reports can be collected in a cost-effective manner. The data were drawn from three studies which achieved fairly high response rates (76% -95%).

In conclusion, Vazire (2005) reports:

> The perplexing pattern of over-reliance on self-reports in the field of personality research seems to be based, at least in part, on the outdated belief among researchers that other methods of assessing personality place a significant burden on the researcher or participants. Unfortunately, this belief has led many researchers to overlook informant reports, which provide rich, valid assessments of personality at minimal cost to the researcher (p. 479).

In conclusion, common method variance continues to be a matter of concern for studies employing a single data collection method for multiple study-variables.

Response-shift bias. This threat to validity arises in the context of pretest and posttest, or repeated observations as in the case of longitudinal studies. While using data through SAQs, researchers assume that the respondents will maintain the same standards for measurement of the construct being assessed across different observation-points (pretest/posttest). It is possible that depending upon personal mood and health, survey-administrative conditions, or for unknown reasons respondents respond in lenient or very strict manner and employ different standards for measurement. Thus, we get varying responses on different occasions. In fact, it is very common for the respondents not to remember their measurement standards of prior responses (Howard & Dailey 1979; Sharpley & Christie, 2007). Accordingly, the comparison of pretest with posttest may have confounding effects of different measurement standards, and may no longer offer valid inferences. Thus, response-shift biases are highly likely to be present as the respondents use varying measurement standards when the same survey instruments are used multiple times on the same subjects.

Howard and Dailey (1979) studied response-shift bias in pre/post self-administered data on an interview skill questionnaire. In addition to pre and post SAQs, *Then* surveys were conducted (along with Post) for which students were asked to report what they felt at the beginning of the workshop. Results indicated significant differences between Pre/Post and Then/Post data.

Howard and Dailey (1979) reported –

> The mean correlation between Pre/Post self-report change and change in judges' Skill ratings was —.06, whereas the mean correlation between Then/Post change ratings and changes in judges' skill ratings was .43. Similar comparisons, substituting behavioral incidents for judges' skill ratings also favored the

> Then/Post Approach (*r*=.33) over the Pre/Post Approach (*r*= -.05) by a similar wide margin (p, 149).

Retrospective data techniques may serve as an effective tool for dealing with response-shift bias. Howard and Dailey (1979) noted – "Self-reported measures of change that used retrospective pretests to remove response-shift bias demonstrated significantly greater validity than measures of change that used traditional self-report pretests" (p. 144). Since then, many researchers have collected retrospective data in an attempt to deal with the potential invalidity of response-shift, especially in pre-test versus post-test designs. For example, Sharpley and Christie (2007) asked 197 breast cancer patients to report their current (about 2 years after diagnosis) anxiety and depression states, and also how they felt at the time of their diagnosis.

Validity Threats: Respondent Characteristics

Del Boca and Noll (2000) suggest various respondent related factors which may affect the validity of their responses in the health sciences. These factors include: respondent's perceived role and response contingencies, anonymity and confidentiality, respondent's personal circumstances (for e.g., physical condition, motivation to respond and psychological and cognitive status), absence/presence of other data sources, and standardization of assessment procedures and settings. Del Boca and Darkes (2003) reviewed literature on the validity of self-reports of alcohol consumption and concluded that self-report validity may be affected by many cognitive factors (e.g., memory, information processing and cognition). Regarding the relationship between demographic variables (e.g., gender, race/ethnicity) and valid responses, they report that research is inconclusive. Critical literature reviews have linked consistency motif (i.e., propensity for respondents to maintain consistency in their responses; Podsakoff et al., 2003; Podsakoff & Organ, 1986), leniency biases (propensity to attribute socially desirable traits, attitudes/behaviors to someone they like than to someone they dislike), mood state (positive or negative emotionality), and acquiescence biases (propensity for respondents to agree [or disagree] with items independent of their content; Podsakoff et al., 2003). Nonetheless, two widely reported threats to self-report validity are social desirability and a joking/lying attitude in respondents.

Social desirability bias. One widely discussed and researched respondent bias is social desirability in SAQs (e.g. Antin & Shaw, 2012; Crowne & Marlowe, 1960; Del Boca & Darkes 2003; Del Boca & Noll, 2000; King & Bruner, 2000; Reynolds, 1982; Podsakoff et al., 2003; Podsakoff & Organ, 1986; Tourangeau & Yan, 2007). Over the years, researchers have expressed concern over distortions introduced into surveys by the respondent's desire to pretend to be socially normal. Podsakoff et al. (2003) stated that social desirability "refers to the tendency of some people to respond to items more as a result of their social acceptability than their true feelings" (p. 882).

One of the most prevalent ways for dealing with social desirability bias is to directly measure the social desirability of respondents. This permits researchers to statistically control for social desirability (as suggested by King and Bruner, 2000). Crowne and Marlowe (1960) developed a social desirability scale consisting of 33 items that revealed high reliability (KR 20 coefficient of 0.88 and test-retest correlation of 0.89). A shorter version (with 13 items) of this scale was developed by Reynolds (1982).

However, this version tends to have relatively lower, but acceptable, reliability (Reynolds, 1982).

In addition, researchers across various social science fields have used the list experiment method to detect social desirability in SAQs. The List experiment technique typically consists of indirect questions used to examine social desirability. It is especially useful when the variable of interest has only three or four response options. In list experiments, samples are randomly divided into k+1 groups (where k= number of response options). One group serves as the control group, which receives the usual type of survey items with all of the k option categories. However, participants in the list experiment condition are given items with a list of k-1 option categories and asked to choose as many choices as they wish. In total, there are k number of list-experiment groups. The difference in mean between the control and the list-experiment group hints at the social desirability bias associated with the omitted response category. For example, Antin and Shaw (2012) examined the relations between social desirability bias and patterns of motivation factors (kill time, fun, sense of purpose, or money) to do work on crowd sourcing services in the US (*N*= 1132) and India (*N*= 898) through internet-based self-reports. Suppose participants in the control condition selected an average of 2.9 out of 4 items (4 motivational factors), while those in treatment condition selected 2.2 out of 3 items. It can be inferred that in the absence of social desirability, 70% (i.e., 2.9 – 2.2) of participants would select the fourth item. The study found that among the US respondents social desirability encouraged over-reporting of all four motivating factors, especially the money-factor; whereas in the Indian sample, social desirability was linked with over-reporting of sense of purpose and under-reporting of killing time and fun.

Invalid respondents (jokesters/ liars/ mischievous respondents). One of the major issues with survey research is the lack of sincerity on the part of the respondents, especially adolescents. There is no way to recruit only sincere respondents. Respondents may lie or choose a certain response option because they find it amusing. Data from such respondents would not be trustworthy.

Many researchers have investigated the effects of non-valid respondents on the validity of SAQs (Cornell, Klein, Konold, & Huang, 2012; Cornell & Looper, 1998; Cornell, Lovegrove, & Baly, 2014; Fan, Miller, Park, Christensen, Grotevant, & Tai, 2006; Cross & Newman-Gonchar, 2004; Robinson-Cimpain, 2014; Rosenblatt and Furlong, 1997).

A large-scale high school survey (*N*= 10,909) conducted by Cornell and Loper (1998) found that nearly 8% of student respondents admitted to giving invalid responses. The invalid responders were likely to report higher rates of fighting, carrying a gun, and using drugs than valid respondents. Cross and Newman-Gonchar (2004) reported that after screening out respondents who gave inconsistent and extreme responses, the estimates of risk behaviors, antisocial behavior, and victim experiences were reduced significantly. Similarly, Cornell et al. (2012) found that student risk behaviors were significantly lower when non-valid respondents were screened out. Also, valid respondents perceived school climate more positively. Fan and colleagues (2006) suggested that jokesters were much more likely to report skipping school, drinking, getting drunk, having physical problems, and being involved in fighting than true respondents.

Dealing with Invalid Responders in SAQs

In order to estimate the bias introduced by the invalid respondents in a study, the first step is their identification. For this purpose, the research literature suggests primarily four mechanisms: bogus pipeline, triangulation of multiple-source data, screening items, and response time data.

Bogus pipeline. It is a technique in which the responders are convinced that their responses can be independently verified by the researchers from some sort of measuring device or other sources of information. This technique is used to counter validity threats in self-reported questionnaires. In their study on smoking behavior among adolescents, Akers, Massey, Clarke, and Lauer (1983) collected saliva samples as a bogus pipeline condition. The data came from a state-wide longitudinal study and were in the form of confidential student questionnaires (Q), anonymous randomized response instruments (R), and saliva specimens (S). The saliva specimens were analyzed only in the first year, while the other study components were kept the same in the second year. In order to study the effect of the bogus pipeline condition, results of year one were compared with those of year two. Findings suggested that the bogus pipeline revealed little effect on the respondent validity in general. Also, self-reports had strong validity because confidential self-reports were consistent with anonymous randomized responses.

Triangulation of multiple-source data. Another approach which researchers often use for identification is that of triangulation through multiple methods. Fan and colleagues (2006) used data (*N*~20,000) from the first wave of the National Longitudinal Study of Adolescent Health (Add Health) study. Add Health data were in the form of student questionnaires administered at school, student interviews at home, and interviewer-assisted parent questionnaires at home. Student questionnaire responses on whether students were adopted, born in United States, and if they were using artificial limb, were matched with interview responses and parent responses at home. Through triangulation of different data sets, Fan et al. (2006) identified valid respondents, inaccurate responders (who may have unintentionally provided an incorrect response due to carelessness or confusion) and jokesters (who might have mischievously provided incorrect responses intentionally). In addition, demographic variables (e.g., gender, age, and ethnicity/race) were also used for supporting identification of inaccurate responders. Results suggested that jokesters were much more likely to report skipping school, drinking, getting drunk, having physical problems, and being involved in fighting than true respondents. However, in most of the SAQ studies it is often the case that all data come from one method and the survey has no screening items. Identification of invalid respondents becomes very challenging in such scenarios.

Screening items. In SAQ data, one of the most challenging tasks is to identify invalid data. However, inclusion of screening items (also called validity items) in the survey makes this job easier. Cornell and Loper (1998) used two items for a validity check: “I am reading this survey carefully” and “I am telling the truth on this survey”. About 8% of students gave a negative response to one or both of these items and were screened out. In a three year longitudinal study, Cornell, Lovegrove, and Baly (2014) examined the validity of responses of 382 middle school students on risk behavior, bullying, and school climate measures. Valid respondents were screened by two items: “I am telling the truth on this survey” and “I am not paying attention to how I answer this survey”. Both items were measured on a 4-point Likert scale (strongly disagree, disagree, agree, strongly agree). Nearly 10% of respondents in each wave reported not telling the truth or not paying attention.

In order to identify inconsistent respondents, Cross and Newman-Gonchar (2004) first selected survey items which were logically nested. For example, a consistent respondent would select “never” and “no”, respectively, for questions, “at what age you belonged to a gang?” and “have you ever belonged to a gang?”. Extreme respondents were the ones who reported “logically implausible, if not impossible” responses (e.g., students who had taken LSD 20+ times in last 30 days; and had drank alcohol on school property 20+ days in last 30 days).

More recently, Robinson-Cimpain (2014) investigated the effects of “mischievous” responders on the validity of outcomes of sexual identity, gender identity, and physical disability. This study applied a four step sensitivity-analysis: 1) identifying items for screeners from all items in SAQ, 2) calculating index values using a screener-indexing approach, 3) examining representation of groups throughout range of index values, and 4) estimating a set of disparities and assessing stability, or incorporating screener-indexing values. Similar to Cross and Newman-Gonchar (2004), this study identified items that were likely to have logically consistent responses. For example, individuals who reported having 2 or more children and being LGBQ were considered for removal from the analytic sample.

Response time data. With the rise of computer-based testing and assessment, psychometricians have new opportunities for screening valid respondents. Wise and

Kong (2005) developed a measure of response time effort (RTE) and hypothesized that unmotivated examinees will answer too quickly even before they have time to read and comprehend the item. This study presented convergent and discriminant validity evidence for RTE and found that the response accuracy of examinees demonstrating rapid-guessing behavior was not significantly higher than chance levels. More recently, Meyer (2010) employed a mixture Rasch model to identify unobserved groups of examinees on a response-time indicator. About 15% of examinees were found to be engaged in rapid-guessing behavior. In conclusion, response time may find applications in validity screening in SAQs.

The selection of items for screening in a post-data collection phase is a challenging task and may not work for surveys that do not have items that explicitly reveal a logical fallacy. Furthermore, even when screening items are identified, it is almost impossible to distinguish between inaccurate respondents and jokesters or mischievous respondents in many cases (Fan et al., 2006). Moreover, if all respondents who report logically flawed responses on a couple of items are considered invalid, the researcher may commit a type I error. In other words, there is a possibility of screening out many respondents who may have selected an invalid response (introduced random error) on the screening item, but responded to other items in a valid manner.

Proposed New Technique

From the literature reviewed so far, it is clear that there is a gap in the methodological literature for identification of invalid respondents when multi-method triangulation of data is not possible, response time data are unavailable, and when screening items are not incorporated. A comprehensive statistical approach to address this problem in SAQs is warranted. The present study proposed a novel three step application of latent profile analysis (LPA) to identify potential invalid respondents (non-serious respondents, jokers, or liars) in three steps:

Step 1: Generate individual-level *response-inconsistency* variable, which gauges the extent to which individual's responses increased or decreased the coefficient alpha for the sample, for each scale in the study.

Step 2: Apply LPA on these response-inconsistency variables. Respondents who exhibited extreme profiles of these response-inconsistency variables (clustering at extreme values of mean) for all study-scales may present invalid groups.

Step 3: Cross-validate the emergent latent profiles in response-inconsistency with responses on other traditional techniques (screening items and response time).

Generating Response-Inconsistency Variables

In measurement theory, reliability is defined as "the degree of test score consistency over many replications of a test or performance task" (Meyer, 2010; p. 4). Accordingly, factors that may prevent perfectly replicated measurements introduce errors (or bias) in estimates. Cronbach, Gleser, Nanda, and Rajaratnam (1972) proposed a comprehensive framework of generalizability theory that takes into consideration multiple sources of errors, and estimates these error variance components separately within a single analysis. From this perspective, when error factors associated with the use of different scale items need to be addressed, internal consistency of the scale is estimated (Cortina, 1993). Internal consistency estimates account for two error-variance components (i.e., variance due to instrument items, and the interaction of persons and items). The most widely reported internal consistency estimates in survey research is alpha coefficient.

Extending Guttman's (1945) reliability framework, Cronbach (1951) defined an index for internal consistency of a scale by the following:

$$\text{Internal consistency, } \alpha = \frac{K}{K-1}\left(1 - \frac{\sum V_{items}}{V_{scale}}\right) \tag{6}$$

Where, K= Total items in a scale;

ΣV_{items} = Summation of item score variances & V_{scale} = Scale Score Variance

From equation 6, we can generate an individual-level "response-inconsistency" variable that has a unique value for each respondent using the following expression:

For each individual, *j* = 1, …, *N*

$$\text{Response-Inconsistency } R_j = \frac{K}{K-1}\left[1 - \frac{\sum_{i=1}^{K}\left[\sigma Y_i^2\left(\frac{n}{n-1}\right) - \frac{\left(Y_{ij} - Mean\, Y_i\right)^2}{n-1}\right]}{\left[\sigma X^2\left(\frac{n}{n-1}\right) - \frac{\left(X_j - Mean\, X\right)^2}{n-1}\right]}\right] \tag{7}$$

Where, K= Total items in a scale,

$\sigma Y_i^{\,2}$ = Score Variance of item Y_i

$\sigma X_j^{\,2}$ = Score Variance of scale *X* for individual *j*

Conceptually, equation (7) generates a response-inconsistency variable for scale *X*, where individual *j* receives an estimate that is the Cronbach's alpha value for the entire sample excluding individual *j*. Response-inconsistency variable gauges the extent to which an individual's responses change alpha coefficient value for the remaining sample for a particular scale. If individual *j* has a very high value on this response-inconsistency variable, it indicates that the Cronbach alpha value for a particular scale *X* would increase if individual *j* is removed from the analysis. In other words, participants with highly inconsistent responses are likely to have higher values on this "response-inconsistency" variable across scales. It is also possible that some invalid respondents consistently select the same rating and have extremely low values on response-inconsistency variables. Accordingly, invalid respondents could form clusters at the extreme values of response-inconsistency variables.

The concept of generating individual-level reliability estimates has been explored earlier by Raju, Price, Oshima, and Nering (2007). Raju et al. (2007) used a binomial error model for estimating individual-level standard errors of measurement for dichotomously scored items. In contrast, the present study conceptualizes individualized response-inconsistency variables differently by modifying the formula of Cronbach's alpha and extends the application to continuous variables.

After computing response-inconsistency variables for all scales of a survey, LPA can be employed on these variables to examine respondent profiles. It was hypothesized that the invalid respondents would have extreme values (too high or too low) on response-inconsistency variables for all scales of survey. LPA is a model-based cluster analysis technique for continuous indicators, and helps detect clustering at the extreme values. The resultant profiles consisting of extreme mean values across all response-inconsistency variables could represent invalid respondents. In addition to this, these latent profiles can be cross-validated through further examination of their association with screening items (if students reported telling truth) and response time (if students reported too fast so that they may not have read the questions carefully) to identify invalid respondents.

Using data from Virginia Secondary School Climate Survey (VSSCS), the present study employed LPA on response-inconsistency variables (based on equation 7) for seven scales (disciplinary structure, academic press, willingness to seek help, respect for student, cognitive engagement, affective engagement, and PTB). The following research questions will be studied:

1. What are the profiles of invalid respondents?
2. What is the association between invalid respondent profiles and respondents who report that they did not tell the truth on the survey or finish survey so quickly that it is unlikely that they provided thoughtful answers?

Patterns of School Climate among High School Students

Policymakers and educational leaders need malleable variables that can be manipulated and used for designing interventions to achieve desirable outcomes in schools. School climate is one of the most important malleable constructs for guiding school reform efforts. Positive school climate has been associated with high student achievement and engagement (Esposito, 1999; Gill, Ashton, & Algina, 2004; Lee, 2012; Wang & Holcombe, 2010), more desirable psychological and behavioral outcomes (Klein, Cornell, & Konold, 2012; Kuperminc, Leadbeater, & Blatt, 2001; Way, Reddy, & Rhodes, 2007), less school violence, disorder and peer victimization (Gregory, Cornell, Fan, Sheras, Shih, & Huang, 2010; Gottfredson, Gottfredson, Payne, & Gottfredson, 2005; Konold et al., 2014; Steffgen, Rechhia, & Viechtbauer, 2013; Stewart, 2003), and less student suspension (Gregory, Cornell, & Fan, 2011).

Given its high policy relevance, many prominent national institutes, including the Centers for Disease Control and Prevention and the Institute for Educational Sciences, have emphasized school climate reforms for promoting school connectedness and dropout prevention (Centers for Disease Control and Prevention, 2009; Dynarski, Clarke, Cobb, Finn, Rumberger, & Smink, 2008). In his policy brief, Cohen (2014) wrote:

> …disciplinary guidelines issued by the U.S. Departments of Education and Justice build on recent support and/or endorsement for school climate reform from the Institute for Educational Sciences, SAMHSA and CDC. And a growing number of State Departments of Education (Connecticut, Georgia, Minnesota and Massachusetts) and large and small districts (from Chicago to Westbrook, Connecticut) are developing school climate

> policies and/or laws that support students, parents/ guardians, school personnel and even community members learning and working together to create safer, more supportive, engaging and flourishing K-12 schools (p. 1).

In order to employ school climate reforms, it is desirable to understand how it can be effectively manipulated. This literature review primarily focused on the construct of school climate, its critical components, and within-school response profiles of high school students. Moreover, the convergent validity of these resultant profiles were explored by examining how these profiles map on to academic and risk behavior outcomes.

Definition and Critical Components

Although there is a general consensus among researchers on its importance, school climate is a challenging construct to measure in a valid manner. Its scope may cover a wide range of organizational, educational, interpersonal, and safety aspects of school experiences and it is often defined in broad terms. For example, Cohen, Mccabe, Michelli, and Pickeral (2009) refer to school climate as "the quality and character of school life". The National School Climate Council (2007) states that school climate includes "norms, values and expectations that support people feeling socially, emotionally and physically safe. People are engaged and respected. Students, families and educators work together to develop, live and contribute to a shared school vision (p. 5)".

Cohen and colleagues (2009) described four comprehensive dimensions of school climate: 1) Safety (e.g. physical safety; social-emotional safety, belief in school rules); 2) Teaching and Learning (e.g. social, emotional, and ethical learning; support for academic learning; support for professional development, leadership); 3) Relationships (e.g. respect for diversity; school connectedness/engagement; school community and collaboration); and 4) Environmental-Structural (e.g. cleanliness, school size, physical surrounding). Thapa et al. (2013) and Zullig, Koopman, Patton, and Ubbes (2010) largely support this description of school climate domains. Nonetheless, it should be noted that school engagement, student engagement, school connectedness, school bonding, school liking, school attachment, and school climate are used by different researchers and can be difficult to disentangle (Libbey, 2004). Given this broad and complex nature of school climate construct, there is relatively little agreement on its critical components (Cohen et al., 2009; Cornell & Mayer, 2010; Thapa, Cohen, Guffey, & Higgins-D'Alessandro, 2013).

The present study operationally defined school climate measure consisting of important and often interrelated dimensions: school structure (disciplinary structure and academic press), support to students (students' willingness to seek help, and respect for students), student engagement (affective and cognitive engagement), and bullying climate (prevalence of teasing and bullying, general victimization in school, and whether student was bullied or bullied others). In total, eight continuous scales and two binary items were used for capturing the construct of school climate (see Table 2). More specific discussion on these defining components is provided as following.

School structure and support. These domains were based on the model of authoritative parenting (Baumrind, 1968) which continues to guide research in child development (Larzelere, Morris, & Harrist, 2013). Literature on authoritative parenting theory indicates that effective parents often apply a combination of strict discipline and emotional support while dealing with their children. Baumrind (1997) argues –" *each extreme contains its germ of truth - the liberal permissive model, that autonomy and self-will are to be cultivated, not punished; the conservative authoritarian model, that discipline, sometimes confrontational or punitive, is required to socialize the child's self-indulgent willfulness (p.g. 321)".*

Table 2
School Climate Definition

Dimensions	Indicators	Studies
School Structure	Disciplinary Structure	Cornell, Shukla, & Konold, 2015; Gregory & Cornell, 2009; Gregory, Cornell & Fan, 2011; Gregory et al., 2010; Konold et al., 2014
	Academic Press	Gregory, Cornell & Fan, 2011; Gill et al., 2004; Lee, 2012; Pellerin, 2005
Support to Students	Respect for Students	Cornell, Shukla, & Konold, 2015; Gregory, Cornell, & Fan, 2011, 2012; Konold et al., 2014
	Willingness to Seek Help	Bandyopadhyay, Cornell, & Konold, 2009; Cornell, Shukla, & Konold, 2015; Eliot, Cornell, Gregory, & Fan, 2010; Gregory, Cornell, & Fan, 2012; Klein, Cornell, & Konold, 2012; Konold et al., 2014
Student Engagement	Cognitive Engagement	Konold et al., 2014; Zullig, Koopman, Patton & Ubbes, 2010
	Affective Engagement	Konold et al., 2014; Zullig et al., 2010
Bullying Climate	Prevalence of Teasing and Bullying	Bandyopadhyay, Cornell, & Konold, 2009; Cornell, Shukla, & Konold, 2015; Klein, Cornell, & Konold, 2012; Mehta, Cornell, Fan, & Gregory, 2013
	General Victimization	Cornell, Shukla, & Konold, 2015

	Two binary items on bullying experience

Lamborn, Mounts, Steinberg, and Dornbusch (1991) found that students whose parents employ a combination of supervision and emotional acceptance were likely to report high levels of psychological competence and low levels of dysfunction. The researchers used two dimensions of parenting (i.e., supervision and emotional acceptance) to classify parents into four groups: authoritative (high supervision and high acceptance), authoritarian (high supervision and low acceptance), indulgent (low supervision and high acceptance), and neglectful (low supervision and low acceptance). Authoritative parenting is also linked with higher school performance and engagement in adolescents (Steinberg, Lamborn, Dornbusch, & Darling, 1992).

Accordingly, authoritative school climate theory hypothesizes two key dimensions of school climate: disciplinary structure and student support (Gregory & Cornell, 2009; Gregory, Cornell, Fan, Sheras, Shih, & Huang, 2010). Disciplinary structure means strict but fair enforcement of school rules and support refers to student perceptions that teachers and other school staff members are supportive, respectful, and willing to help (Konold et al., 2014). Many studies have identified these aspects of school climate; for example, Johnson's (2009) review of 25 studies concluded that "schools with less violence tend to have students who are aware of school rules and believe they are fair" and "have positive relationships with their teachers" (p. 451).

Research has demonstrated linkages between authoritative schools (i.e., schools with high disciplinary structure and high support to students) to lower levels of student aggression toward teachers (Gregory, Cornell, & Fan, 2012) and peer victimization

(Gregory et al., 2010). Drawing data from over 7,300 ninth graders and 2,900 teachers randomly selected from 290 high schools, Gregory and colleagues (2010) found that high structure and high support were significantly associated with less bullying and victimization after controlling for school-size, income, and proportion of ethnic minority in school. Gerlinger and Wo (2014) found that the authoritative school climate approach was more significantly linked with lower bullying victimization in schools than security measures (e.g., more guards, metal detectors, locked entrances and locker checks).

Many nationally representative studies have demonstrated associations between authoritative schools and desirable student outcomes. Drawing data from NELS: 88 (National Education Longitudinal Study), Gill, Ashton, and Algina (2004) employed ASC theory with somewhat different dimensions of school climate: academic press and responsiveness (communal values). Academic press served as an indicator for the demandingness (or expectation) from teachers for academic work. They found that students in responsive schools were likely to have higher engagement and internal control. Pellerin (2005) used the same school climate variables but used data from High School Effectiveness Study. Results suggested that students in authoritative schools reported the highest engagement and lowest dropout rate. Recently, Lee (2012) analyzed US data from the Program for International Student Assessment 2000 (PISA 2000) study and found that teacher support and academic press significantly predicted emotional and behavioral engagement in high school students. Note that the dimension of structure was defined by the measure of academic press in these nationally representative studies.

In sum, school climate dimensions of structure (school discipline and academic press) and teacher support to students have a strong theoretical as well as empirical research base. However, structure and support may not be exhaustive in entirely capturing the school climate construct. Students' engagement (or connectedness) with their school, for example, is widely regarded as a critical factor in academic achievement and school completion.

Student engagement. Engagement refers to a student's cognitive and emotional investment in his or her school (Appleton, Christenson, & Furlong, 2008). Students who are engaged are committed to learning (cognitive engagement) and have feelings of pride and attachment regarding their school (emotional engagement). Christenson, Reschly, and Wylie (2012) concluded that "Student engagement is considered the primary theoretical model for understanding dropout and promoting school completion" (p. v). Engagement is regarded as critical to student motivation, learning, and perseverance to high school graduation (Appleton et al., 2008). Research indicates that students reporting lower levels of engagement are at higher risk of dropping out of high schools (Archambault, Janosz, Morizot, & Pagani, 2009; Janosz, Archambault, Morizot, & Pagani, 2008). It may be desirable to include student engagement as an additional dimension of school climate.

Mehta, Cornell, Fan and Gregory (2013) linked student engagement with bullying climate in high schools. Drawing data from a statewide survey (N=7058) of 9th graders, they found that higher levels of prevalence of teasing and bullying predicted lower commitment to school and less involvement in school activities after controlling for demographic characteristics.

Bullying climate. Prevalence of bullying and peer victimization in schools is one of the widely studied outcomes related to student safety. It is associated with internalizing

(Hasen, Steenberg, Palic, & Elklit, 2012; Leadbeater, Thompson, & Sukhawathanakul, 2014; Reijntjes, Kamphuis, Prinzie, & Telch, 2010), as well as externalizing problems (Reijntjes, Kamphuis, Prinzie, Boelen, Schoot, & Telch, 2011). Furthermore, it often affects children and adolescents in a complex manner. A meta-analysis of 18 longitudinal studies concluded that internalizing problems serve as both antecedents and consequences of peer victimization (Reijntjes et al., 2010). Such reciprocal relationships could be especially problematic for a victim because she or he may get trapped in a vicious cycle of critical psychological problems and increasing peer victimization.

Bullying climate may well be a part of larger pattern of students' involvement in negative behaviors (Pellegrini, Bartini, & Brooks, 1999; Rodkin, Farmer, Pearl, & VanArcher, 2000). Klein, Cornell, and Konold (2012) found significant associations between higher levels of bullying climate with higher rates of risk behaviors (smoking cigarettes and marijuana, drinking alcohol, weapon carrying and physical fighting on school property etc.). While defining bullying climate, it could be useful to capture students' perception of bullying behavior in school as well as their personal experiences of bullying and peer victimization. This is important because significantly more students could have witnessed victimization than self-experienced. For example, Shukla and Wiesner (2014) reported that 81% of students witnessed others being victimized, and only 22% reported being victimized in school. Cornell, Shukla and Konold (2015) differentiated prevalence of teasing and bullying (PTB) scale from general victimization scale in the following words – "[PTB] asked students to report on the prevalence of teasing and bullying they observed in their school… [General Victimization] measure asked students if they have experienced a series of different forms of peer aggression such as fighting and threatening with no reference to bullying". It is important to consider bullying climate consisting of these measures and conceptualize as part of a broader school climate.

Student Reported Patterns of School Climate

Given that school climate is mostly conceptualized as a set of school characteristics, researchers consider school as a level of analysis. However, students (and teachers) are nested within schools, and most school climate surveys are administered to students or teachers. Therefore, many researchers average responses of students within a school for each school climate scale to create school-level scores. This aggregation results in loss of within-school variation in student responses and potential differences among students in their experience of school climate (Konold et al., 2014; Raudenbush, Rowan, & Kang, 1991). Stornes, Bru, and Idsoe (2008) examined motivational climate and found that within-class climate significantly varied more than between-class climate. In his study of the unit of analysis for school climate, Van Horn (2003) provided support for its school-level conceptualization, but remained inconclusive as to whether within-school variations in student responses were mere errors. It is possible that students within a particular school may have varying sets of experiences of their school climate; and there could be meaningful response patterns in student reports within schools.

Mayworm, Sharkey, and Parker (2013) examined student response patterns on ten items measuring school structure and teacher support and obtained four latent classes of school discipline (authoritative, authoritarian, uninvolved and permissive) consistent with

ASC theory. Moreover, their findings suggested that students in authoritative class reported higher levels of perceived school safety.

In a longitudinal study examining student engagement, Janosz and colleagues (2010) discovered seven distinct patterns of middle school students. Three patterns were stable (normative); but four were unstable and varied widely across years. Findings suggested that unstable pathways of engagement were closely associated with higher dropout risks. Overall, there has been very limited investigation of within school patterns of student responses to school climate measures. In fact, there is no study examining student profiles of school climate in a multilevel framework.

The Present Study

In the aggregate, the aforementioned review addressed the critical dimensions of school climate and their importance for desired student outcomes. However, critical gaps remain. We do not know the within-school (student-level) patterns of school climate in high school students, and if the profiles are meaningfully different. We also do not know how student profiles relate to, if at all, with student demographics, academic and risk behavior outcomes. The purpose of this study was to apply multilevel latent class modeling to study the profiles of school climate in high school students. School climate was operationally defined as consisting of eight distinct continuous scales (disciplinary structure, academic press, students' willingness to seek help, respect for students, affective and cognitive engagement, PTB, and general victimization), and two dichotomous items asking students if they were bullied or if they bullied others during the past year. In addition, convergent validity analysis examined relations between the emergent profiles of school climate and student demographics (gender, race/ethnicity, grade-level, and parental education), academic outcomes (self-reported grades and future academic expectations), and risk behaviors.

The present study used secondary data from Konold et al. (under review) that focused on the factor structure of school climate constructs (disciplinary structure, respect for students, willingness to seek help, PTB, cognitive and affective engagement) and their convergent validity within a multilevel framework. However, the current study adopts a person-centered analytic approach (rather than construct-centered approach) and presents new analyses by examining student-level latent profiles of school climate.

The following research questions were studied for this substantive application of LC models:

3. What are the within-school (student-level) response profiles of school climate in high schools?
4. Do the emergent latent profiles of school climate differ on demographic factors (e.g., gender, race/ethnicity, grade-level, and parental educational level at student level)?
5. What is the relation between the emergent latent profiles of school climate and self-reported grades and risk behaviors among high school students?

Chapter 3: Research Methods

This dissertation applied LC modeling techniques to both a measurement problem and a substantive problem. The measurement application of LC modeling (study 1) identified the latent profile of invalid respondents (non-serious respondents, jokers, or liars) from the sample in self-administered questionnaire data. In this novel approach, an individual-level response-inconsistency variable, which gauged the extent to which individual's responses increased or decreased the coefficient alpha for the sample, was generated for each scale in the survey (as per equation 7). The idea was to identify these profiles through application of LC modeling on these response-inconsistency variables. Respondents who exhibited extreme profiles of these response-inconsistency variables (clustering at extreme values of mean) for all study-scales were identified as invalid respondents. The resultant latent profiles were cross-validated through further inspection of their association with screening items and response time data. The following research questions were studied for this measurement application of LC models:

1. What are the profiles of invalid respondents?
2. What is the association between invalid respondent profiles and respondents who do not tell truth and finish survey too fast?

The substantive application of LC modeling (study 2) examined the latent profiles of school climate in high school students in a multilevel framework. A multidimensional school climate measure consisted of eight distinct continuous scales (disciplinary structure, academic press, students' willingness to seek help, respect for students, affective and cognitive engagement, PTB, and general victimization), and two dichotomous items asking students if they were bullied or if they bullied others during the past year. Further, the relations between these emergent latent profiles and academic and risk behavior outcomes were explored to examine the evidence for their convergent validity. The following research questions were studied for this substantive application of LC models:

3. What are the within-school (student-level) response profiles of school climate among high schools?

4. Do the emergent latent profiles of school climate differ on demographic factors (e.g., gender, race/ethnicity, grade-level, and parental educational level at student level)?

5. What is the relation between the emergent latent profiles of school climate and academic outcomes (self-reported grades and academic expectations) and risk behaviors among high school students?

This dissertation used secondary data, but present new analyses for both studies (measurement and substantive). The primary data came from the Virginia Secondary School Climate Survey (VSSCS), which is a part of the state's annual School Safety Audit program (Cornell et al., 2014). Note that the two sets of questions (measurement and substantive) employed somewhat different subsamples. The first set of questions used an unscreened sample that included all respondents because the study purpose was to develop a technique which identifies invalid respondents. Whereas, the second set of questions (substantive) were examined on data screened for invalid responders to obtain more accurate substantive inferences. Invalid respondents were identified based on the new technique proposed in study 1, a screening item (students who reported they were not answering truthfully), and response time technique (students who completed the survey too quickly). In total, the measurement study (study 1) contained 52,012 students from 323 high schools; whereas the substantive study (study 2) sample consisted of 47,631 students from 323 high schools.

Primary Sample and Procedure

The description below was obtained from Cornell et al. (2014). All Virginia public schools which had grades 9, 10, 11, and 12, including schools that did not have a 9^{th} grade, were eligible for the survey. A total of 323 of 324 eligible schools participated in the survey. The school participation rate of 99.7% was achieved with the cooperation of the Virginia Department of Education and the Virginia Department of Criminal Justice Services, who endorsed the study and encouraged participation.

Anonymous online surveys on a Qualtrics platform were administered at public high schools with 9^{th} to 12^{th} grade enrollment in 2014 (between February 2 and May 2). Participating schools were given two options for sampling students: (1) invite all 9^{th}, 10^{th}, 11^{th}, and 12^{th} grade students to take the survey, with a goal of surveying at least 70% of all eligible students (whole grade option); or (2) use a random number list to select at least 25 students from each grade to take the survey (random sample option). Schools were given these options in order to choose a more or less comprehensive assessment of their students.

Schools choosing the random sample option were provided with a random number list along with instructions for selecting students (for more information, see Cornell et al., 2014). All students were eligible to participate except those unable to complete the survey because of limited English proficiency or an intellectual or physical disability. The principal sent an information letter to parents of selected students that explained the purpose of the survey and offered them the option to decline participation.

Student participation rate was defined as the total number of students across all schools who participated in the survey divided by total number invited to take the survey. The overall student participation rate was estimated to be 88.7% (52,012 student participants from a pool of 58,613 students asked to participate). Participation rates were assessed separately for schools choosing the whole grade versus random sampling option. For schools using the whole grade option, the estimated participation rate was 82.9% (21,530 of 25,983). In schools using the random sample option, the estimated participation rate was 93.4% (30,482 of 32,631).

According to reports completed by school principals as part of the survey procedure, the reasons for student non-participation were: the student was absent due to illness (64% for whole grade sampling, and 35% for random sampling schools), a schedule conflict (11% for whole grade sampling, and 19% for random sampling schools), language barrier (5% for whole grade sampling, and 2% for random sampling schools), a student disability (4% for both sampling option schools), the student declined (4% for whole grade sampling, and 19% for random sampling schools), parents declined (3% for whole grade sampling, and 4% for random sampling schools), the student was suspended (3% for both sampling option schools), or some other reasons (such as a computer problem, 5% for whole grade sampling, and 11% for random sampling schools).

A preliminary, unscreened sample total of *N* = 52,012 students (50.3% female) in grades nine (26.4%), ten (25.8%), eleven (24.7%) and twelve (23.1%) from 323 schools completed the survey. This sample will be used for addressing methodological research questions. Based on student self-report, the racial/ethnic breakdown was 57.5% European American, 19.4% African American, 11.1% Hispanic, 4.1% Asian American, 1.8% American Indian or Alaska Native and 1% Native Hawaiian or Pacific Islander, with an additional 16% of students identifying themselves with having more than one race. Approximately 20% of the students reported speaking a language other than English at home. Parent education level was assessed by asking students to choose their parent with the highest educational attainment. Students reported that 19.8% completed post-graduate studies, 23.7% completed a four-year college degree, 15.9% completed a two-year college or technical education degree, 31.2% graduated from high school, and 9.2% did not graduate from high school.

The preliminary sample was screened on three criteria: (1) the time it took students to complete the survey, (2) responses to two validity screening questions, and (3) invalid respondent identified in the first application of LC modelling. As described below, 649 students (1.3% of the sample) who completed the survey in less than 6.07 minutes were excluded because it was judged that they would not have been able to read and carefully answer each question so quickly. An additional 3,336 students (6.4% of the sample) responded to the validity questions that they were not telling the truth on the survey and also were excluded. There were 396 respondents who exhibited invalid patterns in the first study but were not captured by previous two criteria. These cases were also screened out from the analytic sample for the substantive study.

The resulting screened sample consisted of N = 47,631 (51.4% female) participants in ninth (26%), tenth (25.9%), eleventh (24.9%) and twelfth (23.1%) grade. Demographic differences between the unscreened and screened samples were small. The racial/ethnic breakdown was 56.8% European American, 17.9% African American, 10.4% Hispanic, 3.8% Asian American, 1.6% American Indian or Alaska Native, and 0.9% Native Hawaiian or Pacific Islander, with an additional 9.3% of students identifying themselves with having more than one race. Approximately 18.9% reported speaking a language other than English at home. The distribution of parental education was 20% completed post-graduate studies, 24.1% completed a four-year college degree, 16.1% completed a two-year college or technical education degree, 31.2% graduated from high school, and 8.6% did not graduate from high school.

Measures

Disciplinary structure, academic press, respect for students, willingness to seek help, affective and cognitive engagement, and PTB were answered on a four-point Likert-scale (1 = *strongly disagree*, 2 = *disagree*, 3 = *agree*, 4 = *strongly agree*).

Disciplinary structure. A seven-item scale was designed to measure the perceived fairness and strictness of school discipline with items such as “The school rules are fair” and “The school rules are strictly enforced.” The items were derived in part from the Experience of School Rules scale used in the School Crime Supplement to the National Crime Victimization Survey (National Center for Education Statistics, 2005). In the present study, total scores ranged from 7 to 28, with Cronbach’s alpha = 0.78.

Academic press. This scale was newly developed and the items were similar to those developed by Midgley et al. (2000). There are five items and the total scores ranged between 5 and 20. Konold et al. (2014) analyzed the factor structure of this scale and demonstrated significant loadings for all items at both within (student) and between (school) levels. In the present analysis, Cronbach's alpha = 0.72.

Respect for students. This four-item scale was designed to measure the perceived supportiveness of teacher-student relationships with items such as how much they agree that adults in their school "really care about all students" and "want all students to do well". The items were derived in part from the Learning Environment scale (Austin & Duerr, 2005). Scores ranged between 4 and 16, Cronbach's alpha = 0.88.

Willingness to seek help. This scale consisted of four items which came from Bandyopadhyay, Cornell, Konold (2009). Students were asked how much they agree whether they would seek help from an adult in their school if "another student was bullying me." In the present study, total scores ranged from 4 to 16, with Cronbach's alpha = 0.76.

Engagement. This scale was derived from the Commitment to School scale (Thornberry, Lizotte, Krohn, Farnworth, & Jang, 1991) and consisted of two factors, affective engagement and cognitive engagement (for details, see Konold et al., 2014). Mehta, Cornell, Fan, and Gregory (2013) found that a nine-item version of this scale was negatively associated with student reports of the prevalence of teasing and bullying in school. Each factor was measured with three items (e.g., affective item -"I feel like I belong at this school"; cognitive engagement item –"I want to learn as much as I can at school") with total scores ranging from 3 to 12. Cronbach alpha values were .89 and .74 for affective and cognitive engagement, respectively.

Prevalence of teasing and bullying (PTB)**.** This scale asked students about the extent of bullying and teasing they observed at school. Consistent with other measures of bullying (e.g., Juvonen, Nichina, & Graham, 2001; Olweus, 2007), item content was not limited to use of the term "bullying," but included general forms of peer harassment associated with bullying. The items were: (1) Bullying is a problem at this school, (2) Students here often get teased about their clothing or physical appearance, (3) Students here often get put down because of their race or ethnicity, (4) There is a lot of teasing about sexual topics at this school, and (5) Students here get teased or put down about their sexual orientation.

There is strong support for the PTB scale in three previous factor analytic studies (Bandyopadhyay et al., 2009; Klein, Cornell, & Konold, 2012; Konold et al., 2014). The most recent study (Konold et al., 2014) used the same student sample employed in the current study and demonstrated the usefulness of these items for measuring the PTB construct at both the student and school level through multilevel modeling (Dedrick & Greenbaum, 2011). The multilevel confirmatory factor analysis yielded good fit at both the student and school level, revealed all items to yield appreciable loadings for students (> .69) and schools (>.81), and demonstrated reliability estimates of .79 and .88 at the student and school (*M* = 12.62, *SD* = 1.16, Range = 8.42 – 15.62) levels, respectively. In the present study, total scores ranged between 5 and 20, with Cronbach's alpha = 0.85.

General victimization scale consisted of five items designed to ask about general victimization experiences. The items were: "A student stole my personal property", "A student physically attacked, pushed, or hit me", "A student threatened to hurt me", "A student threatened me with a weapon", and "A student said mean or insulting things to me." There were three response options (1 = *no,* 2 = *one time,* 3 = *more than once*). This scale is derived in part from Gottfredson's (1999) Effective Schools Battery and has been used in other studies of peer victimization in schools (Cornell, Gregory, Huang, & Fan, 2013; Klein & Cornell, 2010). In the present study, total scores ranged between 5 and 15 with Cronbach's alpha = 0.76.

Bullying experience. Students were also asked direct questions on bullying victimization and perpetration. Two binary items (0=no, 1=yes) were, "have you been bullied at this school during the past one year?" and "have you bullied others at this school during the past one year?"

Academic outcomes. Two outcomes were self-reported grades and academic expectations. Students were asked, "What grades did you make on your last report card" with seven response choices (1= mostly A's, 2= mostly A's and B's, 3= Mostly B's, 4= Mostly B's and C's, 5= Mostly C's, 6= Mostly C's and D's, 7= Mostly D's and F's). Future academic expectation was captured by – "How far do you expect to go in school"; answer choices were 0= "I do not expect to graduate from high school", 1= "I might or might not graduate from high school", 2= "I expect to graduate from high school", 3= "I expect to graduate from a two-year college or technical school", 4= "I expect to graduate from a four-year college", and 5= "I expect to complete post-graduate studies". Both items were recoded so that higher score reflect higher levels.

Risk behavior. Students responded to five Youth Risk Behavior Study items (Eaton et al., 2008) which were included to measure student risk behaviors in this study.

The items had answer choices ranging from either “0 days” to “20–30 days” or “0 times” to “6 or more times” (CDC, 2010). All scores were recoded to 0= absence of behavior, and 1= presence of risk behavior.

Demographic information. The student survey was used to identify gender (1 = male, 0 = female), grade-levels, dummy variables for ethnicity/race (White as reference group) and parent educational level (proxy for SES). The highest education level achieved by either parent was used as a proxy for socioeconomic status (1= *did not graduate from high school*, 2 = *graduated from a high school*, 3 = *graduated from a two-year college or technical school*, 4 = *graduated from a four-year college*, 5 = *completed post-graduate studies*).

Screening items. The survey included two validity screening items to identify students who admitted that they were not answering truthfully. The first item, “I am telling the truth on this survey,” had four response options: *Strongly Disagree*, *Disagree, Agree,* and *Strongly Agree*. At the end of the survey, the second item was “How many of the questions on this survey did you answer truthfully?” This item had five response options: *All of them*, *All but 1 or 2 of them, Most of them*, *Some of them,* and *Only a few or none of them.* A binary variable of telling truth was created where students answering *Strongly Disagree or Disagree* for the first item and *Some of them* or *Only a few or none of them* for the second item were coded as 0, and other categories were coded as 1. Thus, telling truth served as a screening variable where 0=student not telling truth, and 1=telling truth.

Response time. Survey data also included response time measured as the time when the survey started until it was completed. A binary variable (fast respondents) was created and about 3% of respondents (*N*= 649) who completed survey in less than 6.07 minutes were coded as 1 (for details, see Cornell et al., 2014). Median survey completion time was 14.4 minutes.

Data Analysis Plan

The methodological literature suggests application of multilevel modeling when data have a nested structure. Here, multilevel modeling allows for within-school (student-level) analysis while controlling for between-school variations (Peugh, 2010; Raudenbush & Bryk, 2002). Data management was handled in the STATA 12 statistical package, while the statistical analyses were conducted in M*plus* version 6.1 (Muthen & Muthen, 1998-2014). Latent class models were performed in multilevel framework with students at level-1 and schools at level-2. Standard errors were calculated using a sandwich estimator, while parameter estimations were performed using full information maximum likelihood (FIML) estimator to deal with missing data.

Study 1: Measurement application of LC modeling. LC modeling application was performed in the following steps:

1) Individual-level response-inconsistency variables were generated for each of the seven scales (disciplinary structure, academic press, willingness to seek help, respect for student, cognitive engagement, affective engagement, and PTB):

Response-inconsistency was computed using the following expression,

For each individual, $j = 1, \ldots, N$

Response-inconsistency, R_j =

$$\frac{K}{K-1}\left[1-\frac{\sum_{i=1}^{K}\left[\sigma Y_i^2\left(\frac{n}{n-1}\right)-\frac{\left(Y_{ij}-Mean\,Y_i\right)^2}{n-1}\right]}{\left[\sigma X^2\left(\frac{n}{n-1}\right)-\frac{\left(X_j-Mean\,X\right)^2}{n-1}\right]}\right]$$

Where, K= Total items in a scale,

$\sigma {Y_i}^2$ = Score Variance of item Y_i

$\sigma {X_j}^2$ = Score Variance of scale X for individual *j*

2) Performed LC modeling on Response-inconsistency Variables:

Once the Response-inconsistency variables were calculated for all seven scales, they were treated as indicators in the LC model (Figure 3). Relations between the categorical latent variable and its continuous indicators (response-inconsistency variables) are described by a set of linear regression equations in latent class modeling. For stable parameter estimation, it is important to ensure that the model converges at the global maxima (see description on p. 28). This can be done by specifying different starting values for all parameters which are to be estimated in the model. In M*plus*, maximum likelihood optimization is performed in two stages (see Muthen & Muthen, 1998-2014). 200 random starts for the initial stage and 50 final stage optimizations were specified using the STARTS option in ANALYSIS command (see Appendix A).

In order to identify the best-fitting model that explained the data and optimally described the heterogeneity, several LCA models with increasing number of classes were run and their model fit statistics were compared. The analysis began with a 2-class LC model. Thereafter, the number of classes was gradually increased by one at a time until the model failed to converge or the results no longer made sense. Finally, the model fit statistics (e.g., AIC, BIC, Bootstrapping likelihood ratio-test [BLRT], and classification entropy) were compared to determine the number of classes and identify a model that provided the most meaningful and statistically valid results. Lower AIC and BIC values were considered as a source for evaluating the relative fit of competing models. For BLRT, smaller *p*-values indicate that the k-class model fits better to data than the k-1 class model.

3) Cross Validated Invalid profiles with Traditional Techniques.

After determining the best-fitting model, the screening items (telling truth) and response time were used to examine the convergent validity of this new technic in identifying invalid respondents. The auxiliary (e) function in M*plus* tests for the quality of means across latent classes using a Wald chi-square based on posterior probability-based multiple imputation (for details, see Asparouhov & Muthen, 2014). It should be noted that the variables specified as auxiliary variables are not considered for analysis during the parameter estimation of latent classes. In other words, the auxiliary variable specification does not affect the formation and results of latent classes. Accordingly, telling truth (0=no, 1=yes) and fast response (0=no, 1=yes) were specified as auxiliary variables. In addition, dummy variables for race/ethnicity, gender, academic outcomes and risk behaviors were also included in the model as auxiliary variables to examine how classes differed across these variables.

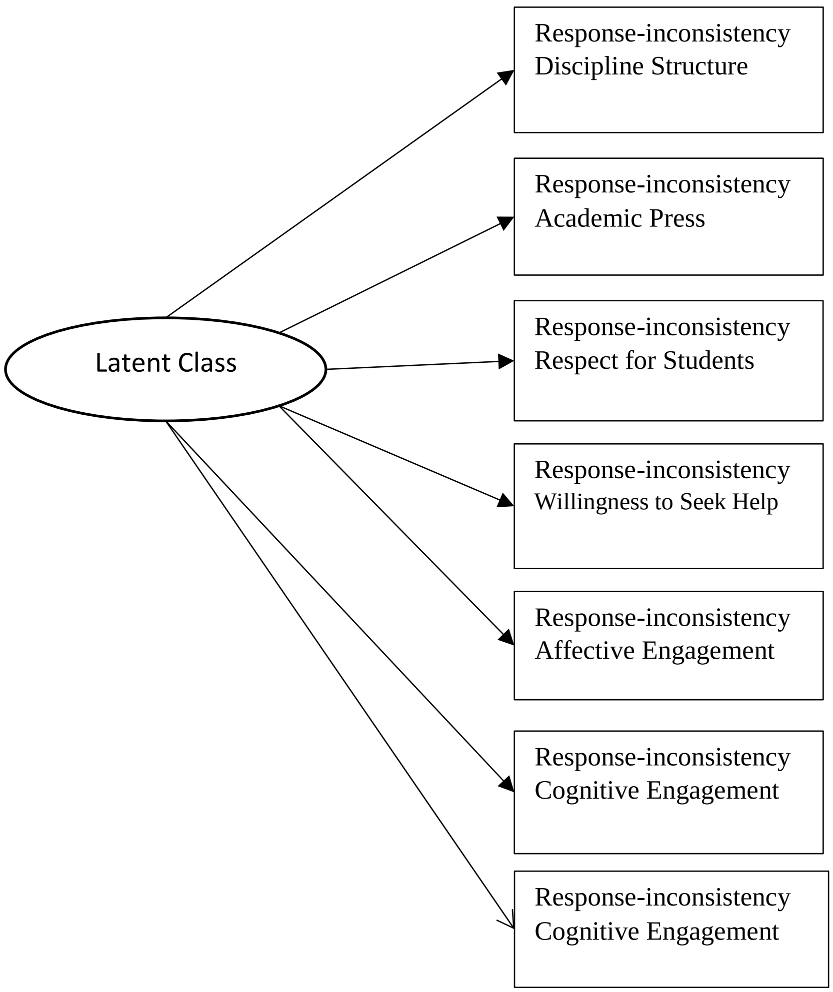

Figure 3. Latent Class Model with Response-inconsistency Indicators

Study 2: Substantive application of LC modeling. Preliminary data analyses included examination of descriptive statistics for all continuous study variables (disciplinary structure, academic press, willingness to seek help, respect for student, cognitive engagement, affective engagement, PTB, and victimization) in the first step.

The data consisted of survey responses from students who were nested within schools. To account for this hierarchical data structure, multilevel latent class modeling was employed. Intraclass correlation values (i.e., ratio of school level variation to student level variation) for eight continuous indicators ranged between 0.03 and 0.12. Note that ICC values in social research usually range between 0.05 and 0.20 (Peugh, 2010). Average cluster-size (number of students in schools on average) was 147.46. Accordingly, design effect values (ranged between 2.61- 17.84) were higher than the recommended cut off value of 2 (Peugh, 2010). In total, multilevel analytic approach was supported by ICCs and design effect values.

In the second step of analysis, latent class modeling was performed on school climate measures which serve as indicators for the categorical latent variable (Figure 4). Relations between the latent variable and its continuous indicators are described by a set of linear regression equations, and between the latent variable and its categorical indicators are described by a set of logistic regression equations in LC models. With ten correlated indicators (eight continuous and two binary variables), the fully free single level model may have 110 unique unknown parameters in a covariance matrix (10 variances + 45 covariances per class) for a two class solution. In order to accommodate latent class modeling within a multilevel framework, covariances were constrained to be zero across classes. This constrained helped resolve model convergence issue, while allowing variances to be freely estimated across classes.

For stable parameter estimation, it is important to ensure that the model converges at the global maxima (see description on p. 28). This can be done by specifying different starting values for all parameters which are to be estimated in the model. In M*plus*, maximum likelihood optimization is performed in two stages (see Muthen & Muthen, 1998-2014). 200 random starts for the initial stage and 50 final stage optimizations were specified using the STARTS option in ANALYSIS command (see Appendix B).

In order to identify the best-fitting model that explained the data and described the heterogeneity optimally, several LCA models with increasing number of classes was run and their model fit statistics were compared. The analysis began with a 2-class LC model. Thereafter, the number of classes was gradually increased by one at a time until the model failed to converge or the results no longer made sense. Finally, the model fit statistics (e.g., AIC, BIC, Bootstrapping likelihood ratio-test [BLRT], and classification entropy) were compared to determine the number of classes and identify a model that provided the most meaningful and statistically valid results. Lower AIC and BIC values were considered as a source for evaluating the relative fit of competing models. For BLRT, smaller *p*-values indicate that the k-class model fits better to data than the k-1 class model.

Finally, convergent validity of the emergent profiles was examined by comparing classes on external variables (demographics, academic outcomes, and risk behaviors). The auxiliary (e) function in M*plus* tests for the quality of means across latent classes using a Wald chi-square based on posterior probability-based multiple imputation (for details, see Asparouhov & Muthen, 2014). It should be noted that the variables specified as auxiliary variables are not considered for analysis during the estimation of latent classes. In other words, the auxiliary variable specification does not affect the formation and results of latent classes. Accordingly, dummy variables for race/ethnicity and grade-level, parental educational-level, gender, self-reported grades, academic expectation, and risk behaviors (weapon carrying physical fight, attempted suicide, drink alcohol, and marijuana use) were included in the model as auxiliary variables to examine how the emergent school climate profiles differed on these variables.

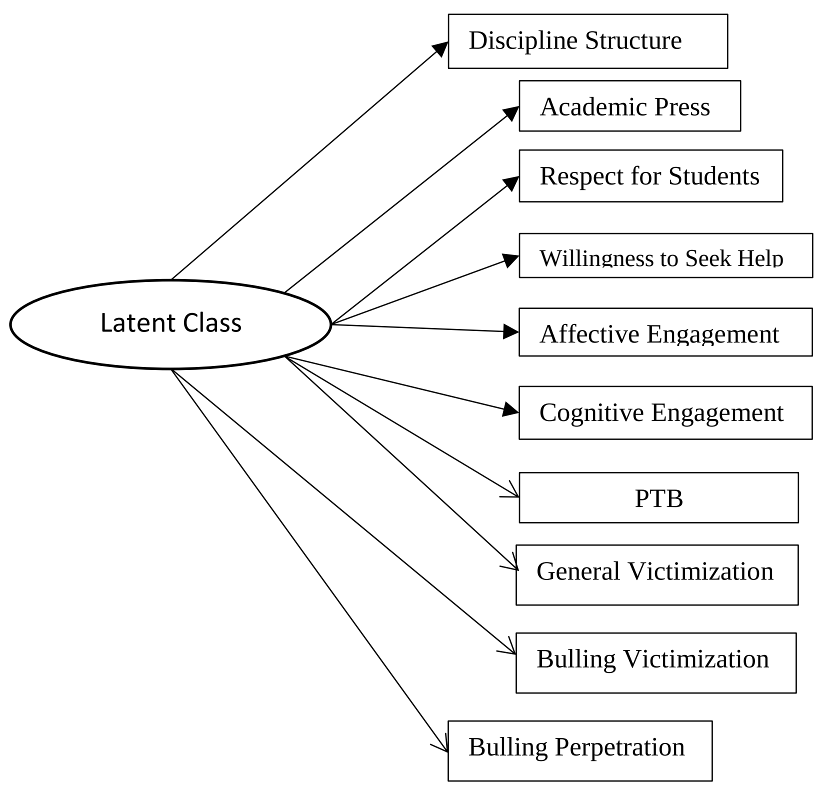

Figure 4. Latent Class Model with School Climate Indicators

Results

Study 1: Measurement problem

Step 1: Creating response-inconsistency variables. Response inconsistency variables were created for school climate scales (structure, academic press, respect for students, willingness to seek help, academic engagement and cognitive engagement) using the formula given in equation 7. Descriptive statistics for these seven variables are presented in Table 3. Figures 5-a through 5-g represents the histograms of RI-structure, RI-academic press, RI-respect students, RI-willingness to seek help, RI-academic engagement, RI-cognitive engagement, and RI-PTB, respectively.

Table 3

Descriptive Statistics for Response Inconsistency Variables

Variables	Mean	Standard Deviation
RI Structure	0.77836	7.25E-06
RI Academic Press	0.72128	1.08E-05
RI Respect Students	0.88081	4.85E-06
RI Willingness to Seek Help	0.75886	1.02E-05
RI Academic Engagement	0.89264	4.90E-06
RI Cognitive Engagement	0.73831	1.18E-05
RI PTB	0.85413	5.84E-06

Note. RI = Response Inconsistency; PTB= Prevalence of Teasing and Bullying

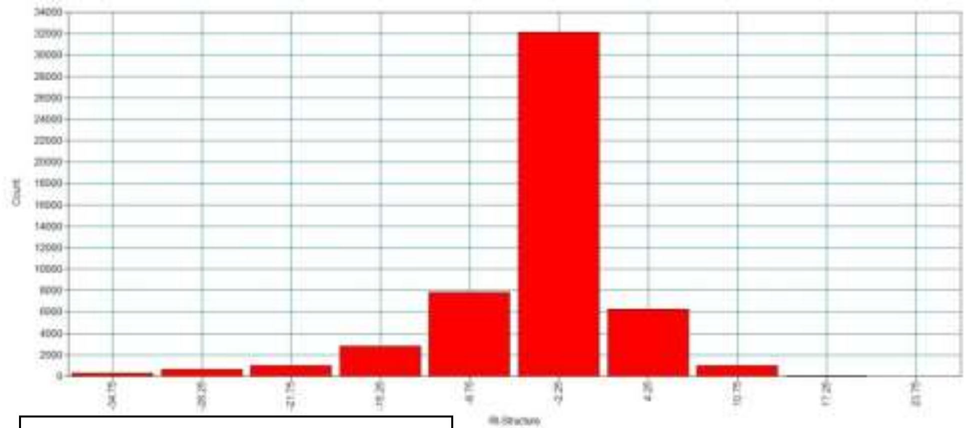

Figure 5a. Histogram of RI-Structure

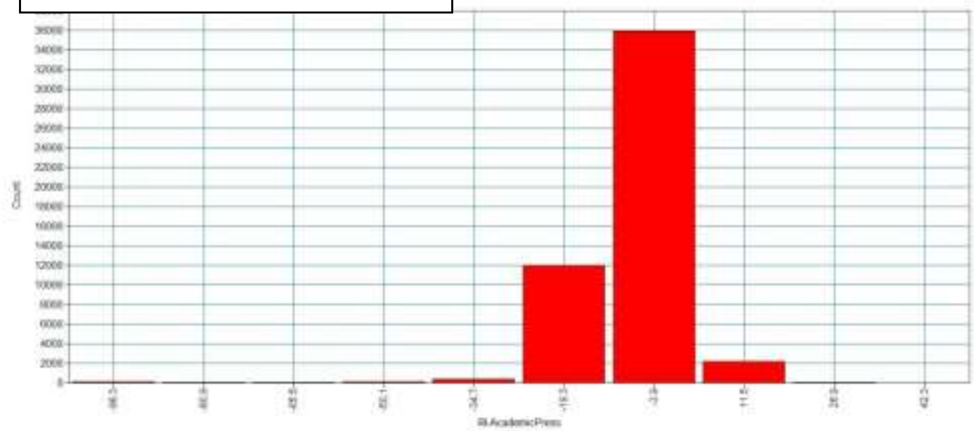

Figure 5b. Histogram of RI-Academic Press

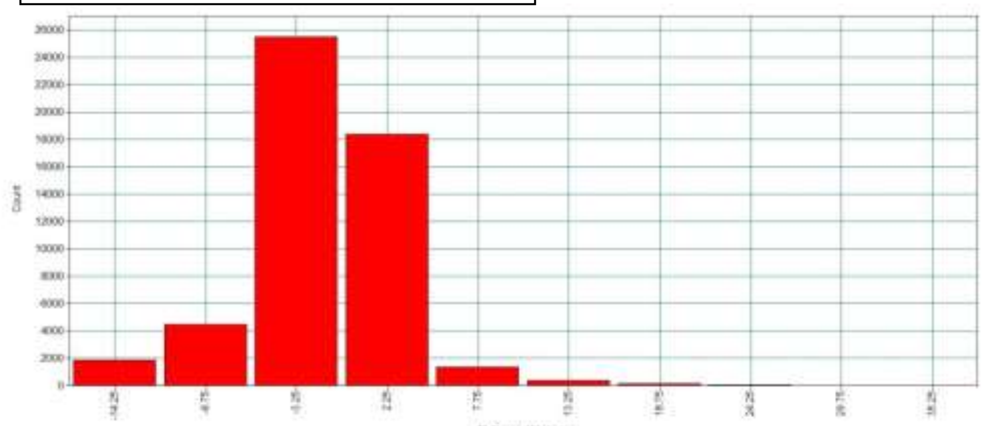

Figure 5c. Histogram of RI-Respect Students

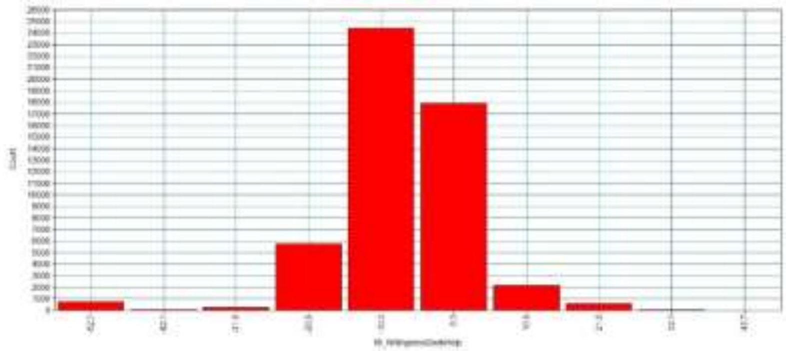

Figure 5d. Histogram of RI-Willingness to seek help

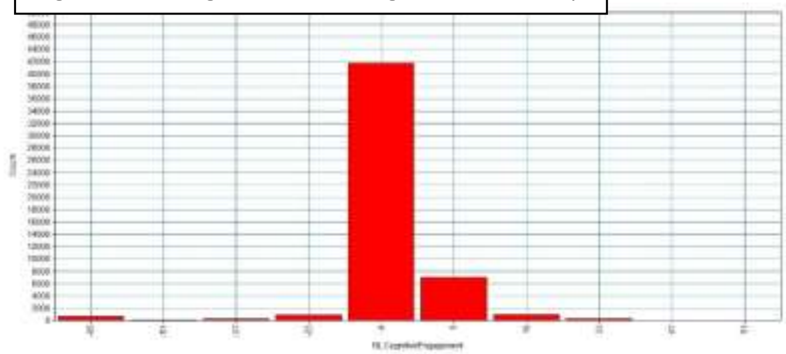

Figure 5e. Histogram of RI-academic engagement

Figure 5f. Histogram of RI-cognitive engagement

Figure 5g. Histogram of RI-PTB

It should be noted that the mean values present the alpha coefficient values for the respective scales. These response inconsistency variables were generated by assigning alpha coefficient value for respondent *j* when respondent *j* was excluded from the sample (Eq. 7). Exclusion of a single case could produce a minute difference in the value of coefficient alpha. Therefore, the resultant values of standard deviations were extremely small (in the range of 10^{-6}; Table 3). For the purpose of latent class modeling, these variables were multiplied by 10^{6} and grand mean centered. Mean centering helped simplify interpretation for the valid respondents across different response inconsistency variables. The dominant clustering around mean of zero would represent valid response profiles, whereas clustering away from zero could suggest invalid profile(s).

Step 2: Application of latent class modeling. Next, latent class modeling was employed on these seven response inconsistency variables. Given that the data structure was nested in this study (students nested within schools), a multilevel approach was adopted. Theoretically, one can argue that the student response pattern (valid or invalid) may not have meaningful associations with their school membership. Accordingly, the intra-class correlation values were low (ranged between .01 - .02). However, the average cluster-size (average students in schools) was 161, and associated design effect was higher (2.92 – 4.52) than the recommended cut-off of 2 for the purpose of ignoring the multilevel analytic approach (Peugh, 2010).

Model fit statistics and information criteria for LC modeling are presented in the Table 4. For the one class model, AIC and BIC values were 2751498 and 2751619, respectively. It was expected that the invalid respondents could form a smaller cluster at extreme values of response-inconsistency variables. Accordingly, variances of these variables were constrained to be equal across groups/classes in order to identify smaller class at the extreme values. Two class model fit the data significantly better (BLRT: χ^2 (8) = 43740, $p < 0.001$) than the one class model (Table 4), with lower values of BIC (2438095) and AIC (2437900). In addition, the entropy value for two class solution was high (.99).

Table 4
Model Fit and Information Criteria

Model	AIC	BIC	LMR-LRT	BLRT	Entropy
1 Class	2751495	2751619	-	-	-
2 Class	2437900	2438095	43242***	χ^2 (8)=43740, $p < 0.001$	0.99
3 Class	2420869	2421135	16853***	χ^2 (8)=17047, $p < 0.001$	0.976

*p< 0.05; **p< 0.01; ***p< 0.001. AIC =Akaike Information Criterion, BIC =Bayesian Information Criterion, LMR LRT =Lo, Mendell, & Rubin – Likelihood-Ratio Test, BLRT =Bootstrapping Likelihood-Ratio Test

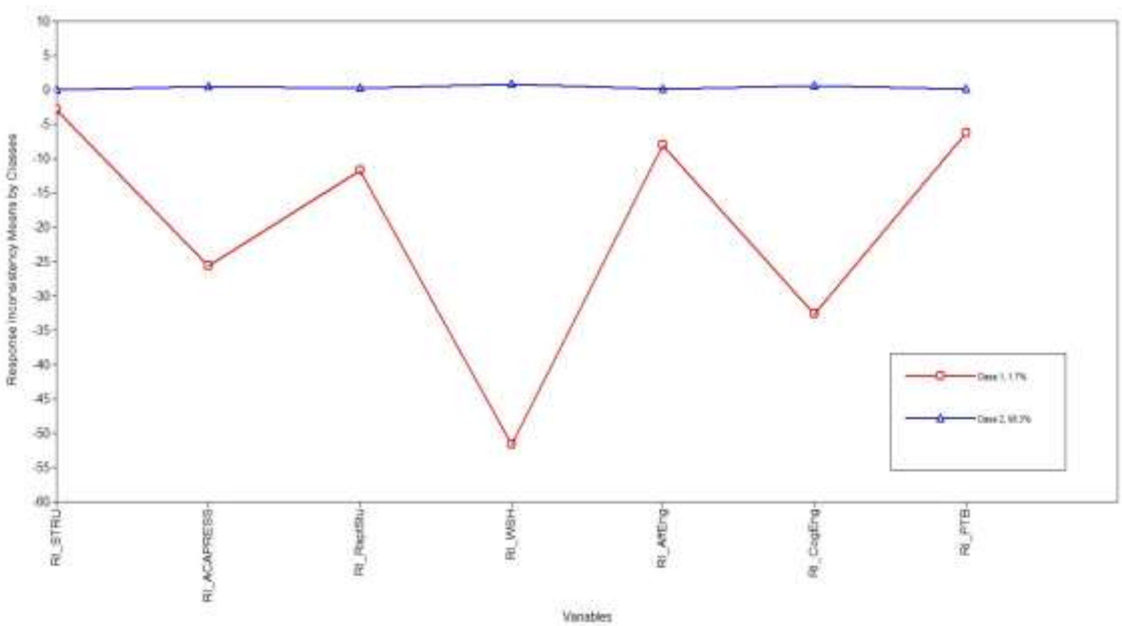

Figure 6. Plot of Centered Means for Response-Inconsistency variables by Classes

A small proportion of students demonstrated an invalid response profile as identified by class 1 (n_1= 878); whereas most students presented with a valid response profile (class 2; n_2= 51,134). Figure 6 represents differences in means for all response inconsistency variables by classes. Note that variables were mean centered prior to analysis; therefore the grand mean values were zero for all response inconsistency variables. As seen in the figure 6, mean values for invalid respondents in class 1 varied significantly from the dominant group of valid respondents in class 2.

Step 3: Cross-validating response profiles. Once the invalid respondent profile was identified, it was compared with the valid respondent profiles through other traditionally employed validity screening techniques (screening item, and response time), demographic variables, and critical student outcomes (peer victimization and health risk behavior).

Table 5

Classification of Sample in Valid/Invalid Groups by Different Techniques

		Response time technique			Screening Item		
		Invalid	valid	Total	Invalid	valid	Total
Proposed Technique	Invalid	151	727	878	403	475	878
	Valid	498	50636	51134	3176	47958	51134
	Total	649	51363	52012	3579	48433	52012
	Sensitivity	23%	-	-	11%	-	-
	Specificity	-	99%	-	-	99%	-
Response time technique	Invalid	-	-	-	243	406	649
	Valid	-	-	-	3336	48027	51363
	Total	-	-	-	3579	48433	52012
	Sensitivity	-	-	-	7%	-	-
	Specificity	-	-	-	-	99%	-

Screening items identified 3,579 (6.88%) respondents, whereas response time technique identified 649 (1.25%) respondents as invalid out of the total sample of 52,012. The invalid respondents identified by the proposed technique were significantly less likely to report that they responded truthfully (54%; 475 out of the total of 878) as compared to the valid respondents (94%), and were more likely (17%; 151 out of the total of 878) to complete the survey so fast that they could not have responded in a valid manner than the valid respondents (<1%; Table 5). Less than 7% of respondents identified as invalid by screening items were fast respondents.

Table 5 also presents sensitivity and specificity of proposed technique with respect to traditional techniques (response time and screening items). Here, sensitivity refers to the ability of a technique to identify invalid respondents when they are identified as invalid by other technique; whereas specificity is the ability of a technique to identify valid respondents when they are identified as valid by other technique (for details see, Lalkhen & McCluskey, 2008). While comparing with response time and screening item techniques, the sensitivity for the proposed technique was 23% and 11%, respectively. Note that sensitivity for response time was even lower (7%) when compared with screening item. For all comparisons, specificity was 99%.

Table 6A presents the mean comparison of invalid and valid groups on these variables. T-tests results do not assume equality of variance for respective variables because the associated variances varied across groups. Importantly, invalid respondents were significantly less likely to report that they responded truthfully (54%) as compared to the valid respondents (94%); $t(884) = -23.54$, $p<.001$. Invalid respondents were more likely (17%) to complete the survey so fast that they could not have responded in a valid manner than the valid respondents (<1%); $t(879) = 12.73$, $p<.001$.

Table 6A

Means by Valid and Invalid Respondent Classes

	Invalid (N= 878)	Valid (N=51,134)	Sig	R^2	t-value	df
Telling Truth+	54%	94%	***	0.385	-23.54	884.07
Responded Fast+	17%	1%	***	0.156	12.73	879.04
Male+	59%	50%	***	0.030	5.33	908.26
Black+	17%	18%		0.001	-1.08	909.28
Hispanic+	31%	11%	***	0.154	12.71	890.82
Asian+	3%	4%		0.003	-1.56	916.34
Multi-race+	12%	9%	*	0.005	2.12	901.95
American Indian+	5%	2%	***	0.016	3.84	889.8
Parental Education	2.67	3.14	***	0.087	-9.24	899.69
Grades	4.09	5.13	***	0.188	-14.38	893.81
Academic Expectation	2.55	3.89	***	0.327	-20.77	887.81
General Victimization	9.56	6.89	***	0.288	18.95	886.07
Bullying victim+	44%	26%	***	0.115	10.83	900.62
Bullied Others+	42%	15%	***	0.221	15.92	892.98
Weapon Carry+	45%	6%	***	0.379	23.17	880.49
Physical Fight+	49%	9%	***	0.382	23.34	881.99
Attempted Suicide+	39%	7%	***	0.305	19.57	874.72
Drink Alcohol+	61%	26%	***	0.338	21.36	894.25
Marijuana use+	58%	16%	***	0.415	25.17	891.46

Note: *p< 0.05; **p< 0.01; ***p< 0.001. "+" indicates dichotomous variable (0= No, 1=Yes)

As shown in the Table 6A, invalid respondents were more likely to report being male, belonging to a racial/ethnic minority (Hispanic, American Indian, multi-racial), and lower parental educational level than valid respondents. Academically, invalid respondents reported lower grades (t[893]= -14.38, p<.001) and academic expectation (t[887]= -20.77, p<.001) than valid respondents. Group differences explained nearly 19% and 33% of variance in self-reported grades and future academic expectation, respectively. 42% of invalid respondents reported being suspended from the school in past one year as compared to 7% of valid respondents.

There were significant group differences for peer victimization and risk behavior reports as well. On average, invalid respondents reported significantly higher scores (9.56) on the general victimization scale when compared to the valid respondents (6.89); and the group difference explained 28.8% of variance in scores. Invalid respondents were more likely to report that they were bullied (44% vs. 26%) and they bullied others (42% vs. 15%) at the school in past year than valid respondents. Overall, invalid respondents reported significantly higher risk behaviors. 45% invalid respondents reported that they carried weapon at school, 61% reported that they drank alcohol, and 58% reported that they used marijuana at least once in past 30 days. Whereas, only 6%, 26%, and 16% of valid respondents reported these behaviors, respectively. Finally, more invalid respondents reported involvement in physical fight at school (49%) and attempting suicide (39%) in past year as compared to valid respondents (9% and 7%, respectively).

Table 6B

Means by Valid and Invalid Respondent Groups based on Screening Item

Screening Item	Invalid (N= 3,579)	Valid (N=48,433)	Sig	R^2	t	df
Male+	65%	49%	***	0.090	20.34	4183
White+	35%	57%	***	0.141	-26.11	4170
Black+	25%	18%	***	0.024	9.86	4000
Hispanic+	21%	11%	***	0.051	14.49	3890
Asian+	4%	4%		0.000	0.19	4116
Multirace+	11%	9%	**	0.002	3.14	4047
AmeInd+	4%	2%	***	0.014	7.27	3808
Parental Education	2.94	3.15	***	0.019	-8.91	4063
Grades	4.33	5.17	***	0.149	-26.3	3961
Academic Expectation	3.23	3.92	***	0.148	-25.85	3854
Victimization	7.37	6.9	***	0.021	9.25	3901
Bullying victim+	25%	26%	*	0.002	-2.58	4153
Bullied Others+	25%	15%	***	0.045	13.56	3944
Weapon Carry+	22%	5%	***	0.138	24.32	3696
Physical Fight+	27%	9%	***	0.136	24.34	3761
Attempted Suicide+	18%	6%	***	0.076	17.5	3748
Drink Alcohol+	39%	26%	***	0.062	16.22	3977
Marijuana use+	36%	15%	***	0.140	25.12	3862

Note: *p< 0.05; **p< 0.01; ***p< 0.001. "+" indicates dichotomous variable (0= No, 1=Yes)

Table 6B and 6C represents characteristics of valid and invalid groups that were identified by screening items and response time variables. As it is shown, the overall pattern in the differences between both groups is presented in the Table 6A. Invalid respondents significantly differed on self-reported academic outcomes (grades and academic expectations) from the valid respondents. The invalid group was likely to consist of more students from ethnic/racial minorities. Most importantly, students in invalid group reported significantly higher rates of bullying, victimization, and risk behaviors than students in the valid group. More students in invalid group reported that they carried weapon at school (screening item - 22% vs. 5%; response time – 30% vs.

6%), were involved in physical fight (screening item - 27% vs. 9%; response time – 31% vs. 10%), attempted suicide (screening item - 18% vs. 6%; response time – 29% vs. 7%), drank alcohol (screening item - 39% vs. 26%; response time – 38% vs. 26%), and used marijuana (screening item - 36% vs. 15%; response time – 37% vs. 16%).

Table 6C

Means by Valid and Invalid Respondent Groups based on Response Time

Response Time	Invalid (N= 649)	Valid (N=51,363)	Sig	R^2	*t*	df
Male+	49%	50%		0.001	0.62	664
White+	43%	55%	***	0.053	6.08	665
Black+	14%	18%	**	0.014	3.04	668
Hispanic+	25%	11%	***	0.094	-8.26	657
Asian+	8%	4%	***	0.022	-3.83	656
Multirace+	7%	9%	*	0.007	2.18	669
AmerInd+	2%	2%		0.000	-0.33	663
Parental Education	2.98	3.14	**	0.011	2.67	660
Grades	5.34	5.11	**	0.017	-3.41	663
Academic Expectation	3.22	3.88	***	0.110	8.98	654
Victimization	7.75	6.92	***	0.045	-5.54	654
Bullying victim+	33%	26%	**	0.018	-3.46	662
Bullied Others+	32%	15%	***	0.107	-8.9	658
Weapon Carry+	30%	6%	***	0.211	-13.16	648
Physical Fight+	31%	10%	***	0.179	-11.89	648
Attempted Suicide+	29%	7%	***	0.187	-12.14	640
Drink Alcohol+	38%	26%	***	0.055	-6.19	657
Marijuana use+	37%	16%	***	0.154	-10.93	656

Note: **p*< 0.05; ***p*< 0.01; ****p*< 0.001. "+" indicates dichotomous variable (0= No, 1=Yes)

Study 2: Substantive Problem

Step 1: Descriptive statistics. The preliminary sample was screened using three techniques: (1) response time (fast respondents), (2) screening item (telling truth), and (3) invalid respondent identified in the measurement application of LC modelling (study 1). Accordingly, the analytic sample for the current study consisted of 47,631 respondents out of the total of 52,012 respondents.

Descriptive statistics are reported for all continuous indicator variables that were used for the latent class modeling (Table 7). Students were also asked direct questions on bullying victimization and perpetration, which served as dichotomous indicators. 26.2% of students reported being bullied, and 14.5% reported that they bullied others (perpetration) at school in past one year. In addition, students responded to five Youth Risk Behavior Study items which were used to examine the convergent validity for the resultant classes. Some of the students reported carrying weapon at school (4.7%), having alcohol at least once (25.2%), and using marijuana (14.6%) during the past 30 days. A minority of students reported being involved in the physical fight at school (8.1%) and attempting suicide (6%) in the past 12 months.

Table 7
Descriptive Statistics

Variables	Mean	Standard Deviation	Min	Max
Structure	18.6	3.74	7	28
Academic Press	15.73	2.22	5	20
Respect for Students	11.28	2.51	4	16
Willingness to Seek Help	12.47	2.09	4	16
Affective Engagement	8.71	2.10	3	12
Cognitive Engagement	9.85	1.63	3	12
PTB	12.72	3.45	5	20

General Victimization	6.86	2.24	5	15

Step 2: LC modeling. To examine the clustering across school climate measures, a set of latent class models were run from one through five classes successively. Model fit statistics (information criteria, BLRT and entropy results) for LC modeling are presented in the Table 8. For the one-class model, AIC and BIC values were 1843327.58 and 1843485.46, respectively. The two-class model indicated significant improvements over the one-class model in fit statistics with lower values of AIC (1763609.6) and BIC (1763934.15), and a significant result for BLRT (χ^2= 79755.96, *p*<.001). As shown in the Table 8, the model with three classes exhibited better fit than that with two latent classes, with lower AIC and BIC values and higher entropy (.83). However, the best-fitting model was the one with four latent classes; AIC= 1701514 and BIC= 1702171. The model with four latent classes had marginally lower entropy (.80), but it demonstrated a better fit with the data than the 3-class model (BLRT: χ^2= 18521, *p*<.001). Unfortunately, the model with five classes failed to converge. In conclusion, a four-class model was selected because it was supported by the model fit statistics and made sense theoretically.

Table 8
Model Fit and Information Criteria

Model	AIC	BIC	BLRT		Entropy
1 Class	1843327.58	1843485.46	-	-	-
2 Classes	1763609.6	1763934.15	chi2= 79755.96	<.001	0.771
3 Classes	1719997.01	1720488.2	chi2=43650.6	<.001	0.829
4 Classes	1701513.6	1702171.44	chi2=18521.41	<0.001	0.801

*p< 0.05; **p< 0.01; ***p< 0.001

Description of four latent classes. The resultant four latent classes were primarily based on the degree of exposure to the positive school climate that the students reported.

These classes were labeled: 1) positive climate class (n_1= 8.911), 2) medium climate- low bullying class (n_2= 13,804), 3) medium climate- high bullying class (n_3= 14,910), and 4) negative climate class (n_4= 10,006). The descriptive statistics for these classes are reported in the Table 9 and Figures 7a and 7b. As the name suggests, students in the positive climate class reported the highest levels of disciplinary structure (*M*= 22.45, *SD*= 2.80), academic press (*M*= 18.12, *SD*= 1.45), respect for students (*M*= 14.07, *SD*= 1.73), willingness to seek help (*M*= 14.76, *SD*= 1.25), academic (*M*= 10.61, *SD*= 1.50) and cognitive engagement (*M*= 11.04, *SD*= 1.07), and lower levels of prevalence of teasing and bullying (*M*= 10.50, *SD*= 3.30) and victimization (*M*= 5.76, *SD*= 1.03) than students in other classes. 12% of students reported being bullied and 4% reported bullying perpetration in the past year in the positive climate class.

The results revealed two medium-climate classes, which differed mainly by the amount of student involvement in bullying behavior. Students in both medium-climate classes exhibited similar levels of structure, academic press, support and engagement. However, only 1% of students in the medium climate-low bullying class reported being bullied and 3% reported bullying others in the past year as compared to 46% and 22%, respectively, in the medium climate-high bullying class (Table 9, Figure 7a, 7b).

Students in the negative climate class reported the lowest levels of structure (*M*= 14.51, *SD*= 3.19), academic press (*M*= 14.13, *SD*= 2.40), respect for students (*M*= 8.51, *SD*= 2.31), willingness to seek help (*M*= 10.75, *SD*= 2.13), academic (*M*= 6.66, *SD*= 2.21) and cognitive engagement (*M*= 8.92, *SD*= 1.95), and highest levels of PTB (*M*= 15.03, *SD*= 3.47), and general victimization (*M*= 8.40, *SD*= 2.82). These students also reported higher rates of being bullied (41%) and bullying perpetration (27%).

Table 9

Descriptive Statistics by School Climate Classes

Variables	Positive Climate n= 8,911	Medium Climate-Low Bullying n=13,804	Medium Climate-High Bullying n= 14,910	Negative Climate n=10,006
	Classes			
Structure	22.45	19.08	18.73	14.51
	(2.80)	(2.41)	(2.66)	(3.19)
Academic Press	18.12	15.18	15.89	14.13
	(1.45)	(1.40)	(1.82)	(2.40)
Respect for Students	14.07	11.38	11.45	8.51
	(1.73)	(1.33)	(1.72)	(2.31)
Willingness to Seek Help	14.76	12.10	12.62	10.75
	(1.25)	(1.33)	(1.70)	(2.13)
Academic Engagement	10.61	8.88	8.85	6.66
	(1.50)	(1.33)	(1.62)	(2.21)
Cognitive Engagement	11.04	9.67	9.94	8.92
	(1.07)	(1.38)	(1.40)	(1.95)
PTB	10.50	11.38	13.58	15.03
	(3.30)	(2.44)	(2.94)	(3.47)
Victimization	5.76	5.17	7.89	8.40
	(1.03)	(0.34)	(1.92)	(2.82)
	Probability for the event happening			
Bullying Victimization	0.12	0.01	0.46	0.41
Bullying Perpetration	0.04	0.03	0.22	0.27

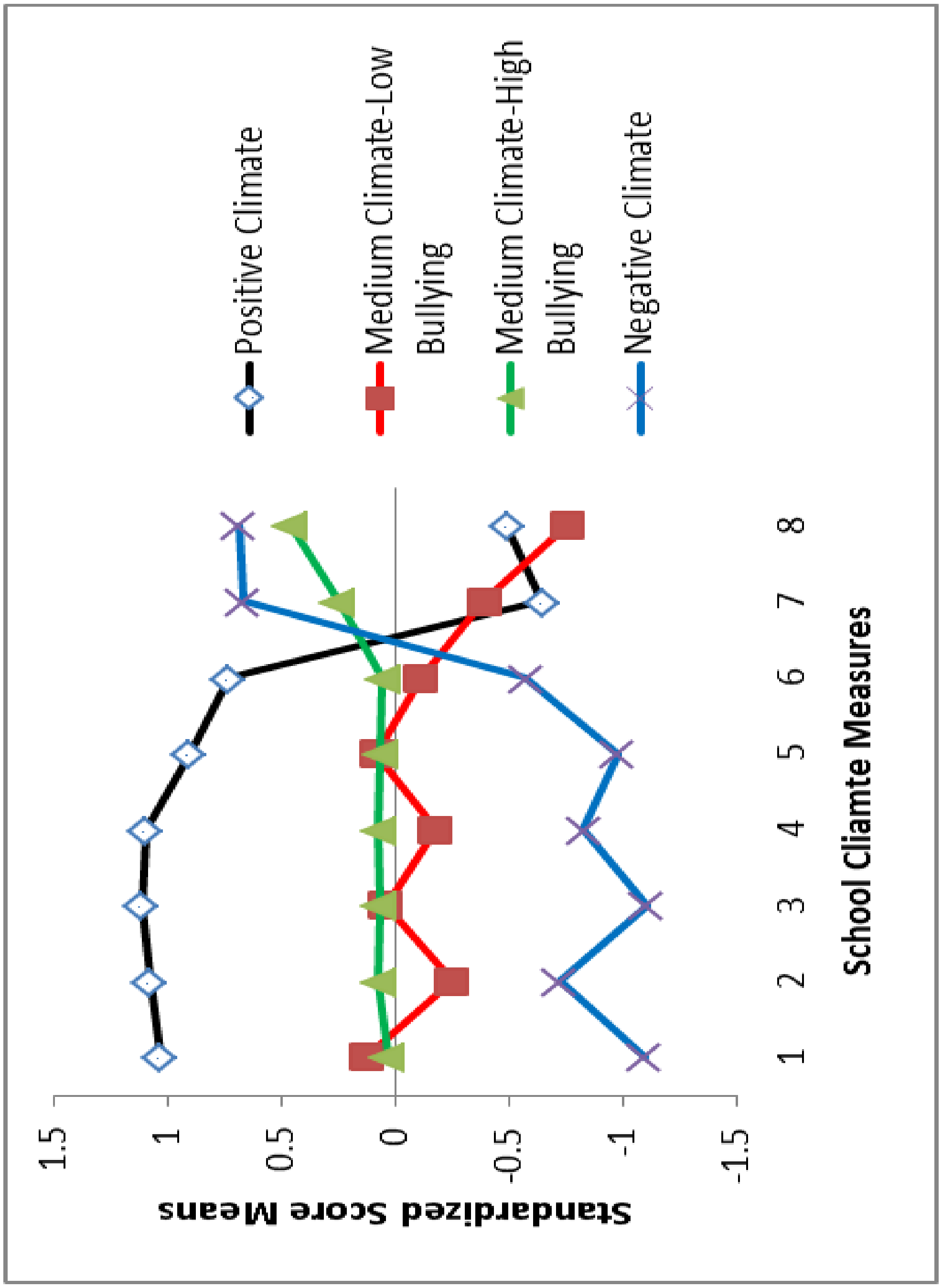

Figure 7a. Means of Continuous School Climate Measures by Classes

Note. School Climate Measures: 1= Structure, 2= Academic Press, 3= Respect for Students, 4= Willingness to Seek Help, 5= Affective Engagement, 6= Cognitive Engagement, 7= PTB, 8= General Victimization

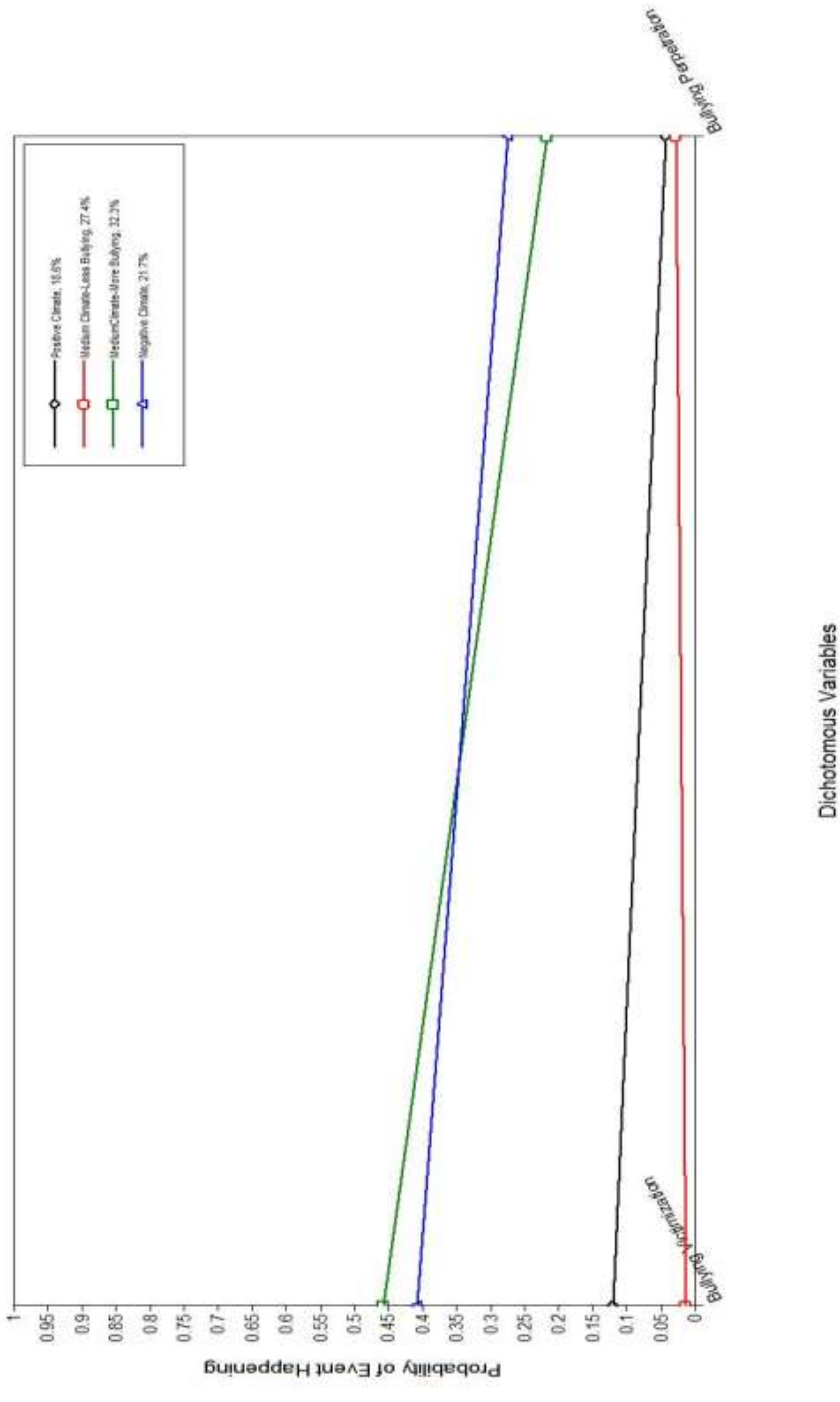

Figure 7b. Probability of Bullying victimization and Bullying perpetration by Classes

Step 3: Cross-validating emergent school climate profiles. Once the latent classes were identified, these classes were compared on external variables (demographic variables, academic outcomes, and risk behaviors). Table 10 presents the mean values (and percentage distribution) of these external variables by classes. In-general, the classes were fairly homogenously distributed by gender groups. The positive and negative climate classes consisted of 48% and 49% of male students, respectively. Latent class of medium climate-less bullying had 50% male students; and medium climate-more bullying class had 48% male students.

Overall, the positive climate class contained more Asian American students (6% vs. 2%) and fewer African American (16% vs. 21%) and multi-race students (7% vs. 12%) as compared to the negative climate classes. Students were quite homogeneously distributed by grade levels across classes. However, there were interesting differences between medium climate-low bullying and medium climate-high bullying classes. The high bullying class contained more 9th graders (28% vs. 24%) and 10th graders (27% vs. 25%), and fewer 11th graders (24% vs. 26%) and 12th graders (21% vs. 25%) than the low bullying class. Average parental education level was significantly higher in the positive climate class (*M*= 3.35) than in the negative climate class (*M*= 2.96; chi^2 [1] = 391.35, p<.001; R^2 value = .02).

There were significant variations in reports on academic outcomes across classes. Students in the positive climate class reported the highest grades (*M*= 5.71) and future academic expectation (*M*= 4.23) and those in negative climate class reported the least (*M*= 4.59 for grades; and *M*= 3.63 for academic expectation; *p*-values< .01 for both outcomes). Effect-sizes associated with the differences between positive and negative climate classes were small; R^2 value = .13 for self-reported grades and .07 for academic expectations.

Table 10.
Means by School Climate Classes

Variables	Positive Climate	Medium Climate Low Bullying	Medium Climate High Bullying	Negative Climate	Chi2	Sig
Male+	0.48	0.50	0.48	0.49	12.33	**
Black+	0.16	0.18	0.17	0.21	56.33	***
Hispanic+	0.10	0.11	0.10	0.11	5.38	
Asian+	0.06	0.04	0.04	0.02	156.6	***
Multirace+	0.07	0.08	0.10	0.12	187.59	***
AmeInd+	0.01	0.02	0.02	0.02	44.98	***
9th grade+	0.27	0.24	0.28	0.25	85.08	***
10th grade+	0.24	0.25	0.27	0.28	16.79	**
11th grade+	0.25	0.26	0.24	0.24	9.57	*
12th grade+	0.24	0.25	0.21	0.23	55.58	***
Parental Education	3.35	3.14	3.19	2.96	331.88	***
Grades	5.71	5.23	5.22	4.59	2132.56	***
Academic Expectation	4.23	3.90	3.99	3.63	1041.52	***
Suspended+	0.002	0.04	0.06	0.13	840.08	***
Weapon Carry+	0.01	0.02	0.04	0.11	901.67	***
Physical Fight+	0.03	0.04	0.09	0.17	1421.89	***
Attempted Suicide+	0.02	0.02	0.07	0.13	1356.83	***
Drink Alcohol+	0.14	0.19	0.27	0.41	2141.73	***
Marijuana use+	0.06	0.10	0.14	0.29	2047	***

Note: *p< 0.05; **p< 0.01; ***p< 0.001. "+" indicates dichotomous variable (0= No, 1=Yes)

Classes significantly differed on prevalence of risk behaviors among students. Negative climate class consisted of students with the highest rates of weapon carrying (11%), involvement in physical fight (17%), suicide attempts (13%), and alcohol (41%)

and marijuana consumption (29%). Alternately, students in positive climate class exhibited the least rates of these behaviors (1%, 3%, 2%, 14%, and 6%, respectively). R^2 values associated with the differences between positive and negative climate classes were: .06 for weapon carrying, .07 for physical fight, .06 for suicide attempt, .10 for alcohol and .13 for marijuana use. Finally, students in medium climate-more bullying reported higher rates of risk behaviors than those in medium climate-less bullying class (4% vs. 2% for weapon carrying; 9% vs. 4% for physical fight; 7% vs. 2% for suicide attempt; 27% vs. 19% for alcohol consumption; and 14% vs. 10% for marijuana usage).

Discussion

The current dissertation presented an overview of the latent class modeling technique and two examples of its application. The first study looked at a measurement issue pertaining to the identification of invalid respondents in self-reported questionnaires (SAQs). The second application examined the latent classes of school climate and their associations with risk behaviors among high school students.

Study 1: Measurement Application

Self-administered questionnaires (SAQs) are widely used for data collection in social sciences. However, respondent characteristics such as joking, lying, and/or responding carelessly pose validity threat to study inferences when SAQs are employed. Traditionally used techniques for identifying such invalid respondents often need prior considerations in study design (e.g., using multiple data sources, additional survey-items, or online survey administration for response time data). This study proposed a new exploratory technique for identifying invalid response profiles in self-reports without such design adjustments.

Invalid respondents were defined as respondents with extremely consistent or inconsistent response profiles which differed from the normative profiles across seven survey scales (disciplinary structure, academic press, support to students, student’s willingness to seek help, academic engagement, cognitive engagement, and prevalence of teasing and bullying). The proposed technique was conducted in three steps: 1) creation of a response-inconsistency variable that gauged the extent to which the individual’s responses increased or decreased the coefficient alpha for the sample for each scale, 2) application of latent class modeling on these variables to examine the clustering at extreme values of the response-inconsistency variables, and 3) cross-validation of cases identified as invalid with traditionally used techniques like screening items and response time data.

Application of latent class modeling on response-inconsistency variables revealed a meaningful two-class solution. One latent class consisted of a minority of respondents (*N*= 878; 1.7%) who demonstrated extremely low mean values across seven response-inconsistency measures and was labeled as an invalid class. Using the present technique, the class with extreme mean values primarily captures the cluster of participants who respond in an arbitrary manner or respond highly consistently in comparison to the normative respondent group. It should be noted that the survey administration was standardized for all respondents and the measurement model for the VSSCS survey has been well established (Konold et al., under review). Therefore, an atypical response profile could reflect atypical respondent characteristics. The second class contained most of the respondents (*N*= 51,134; 98.3%) with the mean values clustered at zero across seven response-inconsistency variables, which were mean-centered for analytic purposes.

The percentage of invalid respondents identified based on this technique (1.7%) was similar to the percentage of fast respondents (1.25%), but lower than those identified based on screening items (6.88%). The proportion of invalid respondents in the sample often depends upon the technique being used. Cornell et al. (2015) reported a relatively smaller percentage of invalid respondents based on response time data (0.7%), but a higher percentage based on screening items (6.4%). In fact, the screening item technique tends to identify relatively more (6% - 12%) respondents as invalid (Cornell et al, 2015;

Cornell, Lovegrove, & Baly, 2014) as compared to other techniques. For example, Fan and colleagues (2006) identified 2.5% of respondents as inaccurate using a triangulation method where student responses were compared with data from parent questionnaires. It is important to be mindful of these variations in employing techniques for identifying invalid respondents.

Cross-validation with traditionally used techniques. The invalid respondents identified by the proposed technique were significantly less likely to report that they responded truthfully (54%) as compared to the valid respondents (94%), and were more likely (17%) to complete the survey so fast that they could not have responded in a valid manner than the valid respondents (<1%; Table 5). These results suggested a significant overlap of this new technique in identification of invalid respondents with other techniques. Interestingly, out of 878 invalid respondents, only 340 respondents were uniquely identified by the proposed technique. Interestingly, there was relatively little overlap between traditionally used techniques of screening item and response time. Less than 7% of respondents identified as invalid by screening items were fast respondents (Table 5). It is possible that response time data helps capture non-serious respondents who perceive survey to be burdensome and just want to finish it, whereas the screening item method detects a larger group of jokesters or liars who are willing to admit they did not tell the truth. One plausible speculation would be that fast respondents could exhibit a highly consistent response pattern (selecting the same category for many items), whereas liars or jokesters could demonstrate more inconsistent and extreme patterns. Accordingly, the proposed new technique significantly overlapped with both the screening item and response time techniques. However, it is not possible to know the intent of respondents from their response pattern. Hence, the identification of invalid respondents remains exploratory in nature.

Demographic examination of the valid and invalid groups indicated that there were more boys and students from racial/ethnic minorities in the invalid group. This finding is consistent with previous research which found that self-reports from boys or non-white students may have relatively higher validity problems than girls or white students, respectively (e.g., Cornell, Lovegrove, & Baly, 2014; Kuncel et al., 2005). It is difficult to speculate on why boys and students from minority groups tend to respond in an invalid manner more. The higher proportion of Hispanic students in the invalid group based on the response time cutoff may have to do with language difficulties (finish survey quickly without reading carefully), which needs further investigation. However, it is possible that invalid respondents find it appealing to claim that they belong to a minority group. For example, 5% of invalid respondents identified by the proposed technique specified they were American Indians, when state enrollment records indicate that <0.01% of Virginia high school students belong to this group.

This study also examined how valid and invalid respondents identified by this technique differed across important academic and risk behavior outcomes. Research suggests that the invalid respondents identified by screening items and/or triangulation techniques are significantly likely to report unusually high levels of risk behaviors and poor academic outcomes as compared to valid respondents (Cornell, Klein, Konold, & Huang, 2012; Cornell & Looper, 1998; Cornell, Lovegrove, & Baly, 2014; Cross & Newman-Gonchar, 2004; Fan et al., 2006). Cornell et al. (2012) found that students who did not agree to telling truth on the survey reported higher rates of weapon carrying (21%

vs. 5%) and involvement in physical fights (24% vs. 11%) at school, alcohol consumption (31% vs. 18%), and marijuana use (30% vs. 12%) than those who agreed. Similarly, invalid respondents identified by the new technique proposed in this study demonstrated significantly higher rates on these outcomes as compared to the valid respondents.

Moreover, comparison between invalid and valid respondents identified by the three techniques (proposed technique, screening item, and response time) found a similar pattern consistent with the above mentioned prior research (see Tables 6A, 6B and 6C). For involvement with bullying and risk behaviors, all three techniques found similar rates for valid respondents (bully victim – 26%; bullied others – 15%; weapon carrying – 5%-6%; physical fight – 9%-10%; attempted suicide – 6%-7%; drank alcohol – 26%; used marijuana – 15%-16%). For invalid respondents, the prevalence rates differed across techniques, but were significantly higher than for valid respondents (bully victim – 25%-44%; bullied others – 25%-42%; weapon carrying – 22%-45%; physical fight – 27%-49%; attempted suicide – 18%-39%; drank alcohol – 39-61%; used marijuana – 36%-58%). It is possible that respondents, who demonstrate extreme levels of consistency or inconsistency in responses, acknowledge that they were not telling the truth, and/or respond so fast that they may not have read all the questions carefully, do not take the survey seriously. They could be joking and may find it amusing to report higher prevalence of negative behaviors like bullying and risk behaviors in their responses (Cornell, Lovegrove, & Baly, 2014).

These findings underscore the importance of screening out invalid respondents from the study sample prior to statistical analysis for improving the internally validity of inferences. Invalid respondents could introduce bias and inflate the rates of negative outcomes (e.g., risk behaviors, victimization and bullying rates, low achievement) for the overall sample.

Significance of measurement application. In conclusion, this study introduced a new method for exploring profiles of invalid respondents in self-reports. Self-administered questionnaires are one of the most widely used methods across various fields of social sciences. This novel technique is intended to help deal with validity threats due to respondent characteristics such as joking, lying, and/or responding carelessly. Traditionally used techniques need either multiple data sources (e.g., triangulation method), additional items planned in-advance (e.g., screening items), or online survey administration (e.g., response time) for creating flags for invalid respondents. However, this newly proposed technique can be used even when none of the other techniques are feasible. Nonetheless, it should be noted that there were several respondents who were uniquely identified by different techniques. Therefore, provisioning for multiple techniques (screening items, multiple data sources, response time, and present statistical approach) for identifying and screening out invalid respondents is recommended. Researchers across different fields of social sciences may find this technique useful in screening out invalid respondents in order to obtain more accurate results.

Limitations and future research. The proposed technique is exploratory in nature. It may be possible that the emergent profiles with extreme means on response-inconsistency measures may still consist of valid respondents. Future research may explore possibilities for generating individual-level reliability coefficients using bootstrapping procedure and employ LPA of these reliability variables. It will be necessary to examine whether the results obtained using the current method can be replicated. In addition, a highly desirable next step for studying this technique would be a simulation study. A mixture of population distributions with three different coefficient alpha values (e.g., .20, .70, and .95 for a five point Likert scale) can be created. The sample size of each group, number of scales and number of items within scales could serve as different experimental conditions. The proposed technique can then be applied for these conditions. Such a study could reveal how effectively this technique identifies random response patterns (respondents belonging to group with alpha=.20) and highly consistent response patterns (respondents belonging to group with alpha=.95) from the normative response pattern (respondents belonging to alpha coefficient= .70 group).

Study 2: Substantive Application

The present study adopted a person-centered approach to analysis by examining latent profiles of a multidimensional school climate measure among high school students. In order to achieve more accurate inferences, invalid respondents were identified by multiple techniques (new technique proposed in study 1, screening item, and response time technique) and were screened out from the sample prior to analysis for this study. There were some major findings of this substantive application of latent class modeling.

Four meaningfully different within-school student profiles were revealed: positive climate class (8,911; 18.71%), medium climate-low bullying class (13,804; 28.98%), medium climate-high bullying class (14,910; 31.30%), and negative climate class (10,006; 21.01%). As their names suggest, the formation of classes was primarily based on the degree of constituent characteristics of school climate. Students reporting higher levels of disciplinary structure, academic press, respect for students, willingness to seek help, academic engagement and cognitive engagement also reported lower levels of PTB, general victimization, and probability of being bullied and bullying others. These findings are consistent with a large body of prior research on authoritative school climate (Cornell, et al., 2015; Eliot et al., 2010; Gill et al., 2004; Gregory et al., 2010, 2011; Konold et al., 2014, under review; Lee, 2012; Pellerin, 2005). Students who experience higher levels of academic press and support from teachers are likely to be more engaged with their school (Gill et al., 2004; Pellerin, 2005). In addition, fair disciplinary structure and supportive environment may encourage students to seek help from adults in the schools. This in return may help reduce prevalence of bullying and teasing (Bandyopadhyay et al., 2009, Eliot et al., 2010). Recently, Konld et al. (under review) found significant associations between disciplinary structure, academic press, willingness to seek help, respect for students, and prevalence of teasing and bullying at both student and school levels. In total, different school climate measures often exhibit significant correlations, which explain why students reporting higher structure, support, engagement and lower PTB cluster together.

The second research question dealt with variations across latent classes on demographic characteristics. Overall, the classes demonstrated fairly homogenous demographic patterns. Although positive climate class consisted of fewer African American students (16% vs. 21%) and had higher mean parental education (3.35 vs. 2.96) than negative climate class, the associated effect size values were small ($R^2 <.02$). Nonetheless, literature suggests that African American students are less willing to seek help in schools (Eliot et al., 2010), and feel less safe at school than other groups (Lacoe, 2015). It is possible that due to this relative sense of insecurity, students perceive lower disciplinary structure and are less engaged with the school.

Another interesting finding was the difference in bullying involvement between two medium climate classes. For medium climate, the class with higher bullying tended to have lower grade students and slightly more females than the one with lower bullying. Literature does suggest a pattern of reduction in peer victimization rates from middle school to later high school grades (Nansel, Overpeck, Pilla, Ruan, Simons-Morton, & Scheidt, 2001). A US-wide survey (Robers et al., 2013) on school crime and safety suggested a gradual decrease in student reported bullying victimization from grade six (30%) to grade twelve (12%). In addition, it is possible that the older students may have been gradually desensitized and/or may not perceive lower levels of bullying as victimization. This could reduce their rates of bullying reports.

In addition, research suggests that girls tend to experience more forms of victimization than boys, with the exception of physical bullying, which occurs more commonly among boys (Bradshaw, Waasdorp, & Johnson, 2014). Accordingly, the overall bullying and victimization reports from girls could be higher than boys.

To answer the third research question of this substantive application of LC modeling, relations between the resultant profiles and academic outcomes and risk behaviors were studied. A major finding was that these resultant profiles demonstrated sound convergent validity; for example, students in the positive climate class reported significantly higher grades (R^2 = .13) and future academic expectations (R^2= .07). Many studies have linked students' academic outcomes with various dimensions of school climate (Bear, Gaskins, Blank, & Chen, 2011; Hoy & Hannum, 1997). For example, Lee (2012) revealed that students who experience a more supportive climate in school perform better academically. Findings from Wang and Holcombe (2010) revealed significant associations between support from teachers, school engagement, and student grades. Goodenow (1993) found that support from teachers explained more than one third of the variance in students' expectations for academic success. Clearly, a positive climate that provides fair disciplinary structure and strong support to students is likely to be conducive for learning and students are likely to succeed academically (Bear et al., 2011).

In addition, there were significant differences between students in the positive climate class and negative climate class on risk behaviors (R^2 values ranged between .06 and .13). These results were well in line with the prior work relating lower risk behaviors with lower prevalence of bullying and higher willingness to seek help from teachers among high school students (Klein et al., 2012). These findings are also consistent with the literature arguing for positive school climate as an important prohibitive and protective factor against risk behaviors in adolescents (Resnick et al., 1997).

Study Implications. The present study underscores the importance of positive school climate. Through a person-centered analytic approach, it revealed within-school profiles of school climate that were operationally defined by dimensions of structure (disciplinary structure, and academic press), support (respect for students, and willingness to seek help), engagement (cognitive and affective), bullying climate (prevalence of teasing and bullying, general victimization, if student was bullied, and if student bullied others). It is important for teachers, principals, and other staff members to be mindful of the within-school variation in perceived school climate. They may see their approaches to school discipline, academic expectation, and supportive climate uniform across all students, but the students do not necessarily perceive it the same way. In fact, students exhibiting negative outcomes (academic, risk behavior, bullying involvement) are likely to perceive lower levels of structure and support.

Although the clustering of students based on degrees of school climate seems intuitive, it poses a fundamental question: why do students within a school perceive school climate differently? This opens up new avenues for examining school climate. In the field of curriculum and instruction, differentiated instruction serves as a framework for catering to various clusters of students with varying academic abilities (for details, see Tomlinson, 1995). Similarly, for catering to various clusters of students perceiving varying degrees of school climate, should there be differentiated school climate interventions? Using the authoritative school climate theory, should adults in school (teachers, administrators, and other staff) provide differentiated disciplinary structure and supportive environment for reaching out to those students who perceive negative climate? Schools tend to take up highly standardized approaches for creating and maintaining systems dealing with school discipline and student support. It is important for the adults in the schools to be mindful of these meaningful within-school clusters of perceived climate across students. One size fits all and unnecessary harsh disciplinary approach could further exacerbate student outcomes, especially for at-risk groups (Losel, 2011). For effective school climate, it may be desirable to examine a possibility for a graded supportive environment, especially by providing higher support and clearer (and more consistent) disciplinary structure for students exhibiting adverse academic and behavioral outcomes. Becker and Luthar (2002) aptly stated – "Disadvantaged students should benefit greatly from access to supportive teachers within the context of a rich and challenging curriculum" (p 202). Using this knowledge of within-school profiles of school climate, school psychologists may like to assist administrators and teachers with conceptualizing a graded approach where at-risk students are provided higher levels of structure and support.

Limitations and Future Research. The cross-sectional design of the present study limits any causal inferential claim. Relations between study variables are correlational. Accordingly, the directionality between variables is speculative in nature. It could be possible that students with lower grades do not like their school and report a negative climate.

In addition, all the study variables were drawn from self-reports from students. Therefore, there could be some bias in inferences due to the common method variance. Common method variance, which is “attributable to the measurement method rather than to the constructs the measures represent” (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), is considered an issue of concern especially when data are from a single source. The reader should be mindful of these limitations before developing any theoretical assumptions and/or working hypothesis between variables based on the findings presented in this study.

However, the study consisted of a large state-wide sample (N= 47,631) and examined latent profiles of student-reports of a comprehensive multidimensional measure of school climate. Clearly, more research is needed for deepening our understanding of school climate and how, why, and when these clusters form within schools. Future study may focus on similar analyses in a longitudinal framework, where one can examine whether change in structure and support over time is linked with shift in membership of students from the negative to positive school climate class, and vice-versa. It could also be interesting to study whether schools with higher means and smaller within-school variation on structure and support dimensions (authoritative equitable schools) do better than those with lower means and larger within-school variation on important student outcomes of engagement, achievement, and risk behaviors.

References

Abar, B., & Loken, E. (2012). Consequences of fitting nonidentified latent class models. *Structural Equation Modeling: A Multidisciplinary Journal, 19*, 1-15.

Agresti, A. (2002). *Categorical data analysis* (Vol. 359). John Wiley & Sons.

Aitkin, M., Anderson, D., & Hinde, J. (1981). Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society*. *Series A (General), 144*, 419-461.

Akers, R. L., Massey, J., Clarke, W., & Lauer, R. M. (1983). Are self-reports of adolescent deviance valid? Biochemical measures, randomized response, and the bogus pipeline in smoking behavior. *Social Forces, 62,* 234-251.

Antin, J., & Shaw, A. (2012, May). Social desirability bias and self-reports of motivation: A study of amazon mechanical turk in the US and India. In Proceedings of the SIGCHI Conference on *Human Factors in Computing Systems* (pp. 2925-2934). ACM.

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, *45,* 369-386.

Archambault, I., Janosz, M., Morizot, J., & Pagani, L. (2009). Adolescent behavioral, affective, and cognitive engagement in school: Relationship to dropout. *Journal of School Health*, *79,* 408-415.

Asparouhov, T., & Muthén, B. (2014). Structural equation models and mixture models with continuous non-normal skewed distributions. *Mplus Web Note, 19,* 1-49.

Austin, G., & Duerr, M. (2005). *Guidebook for the California Healthy Kids Survey. Part III: School Climate Survey for teachers and other staff 2005–2006 edition*. Retrieved from http://eric.ed.gov/?id=ED486328

Babbie, E. R. (2013). *The practice of social research*. Cengage Learning.

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.

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: a unified approach* (Vol. 899). John Wiley & Sons.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: potential problems and promising opportunities. *Psychological Methods*, *9*, 3.

Baumrind, D. (1968). Authoritarian vs. authoritative parental control, *Adolescence, 3*, 255-272.

Baumrind, D. (1997). The discipline encounter: Contemporary issues. *Aggression and Violent Behavior, 2*, 321-335.

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.

Becker, B. E., & Luthar, S. S. (2002). Social-emotional factors affecting achievement outcomes among disadvantaged students: Closing the achievement gap. *Educational Psychologist*, *37*, 197-214.

Bradshaw, C. P., Koth, C. W., Thornton, L. A., & Leaf, P. J. (2009). Altering school climate through school-wide Positive Behavioral Interventions and Supports: Findings from a group-randomized effectiveness trial. *Prevention science*, *10*, 100-115.

Bradshaw, C. P., Waasdorp, T. E., & Johnson, S. L. (2014). Overlapping verbal, relational, physical, and electronic forms of bullying in adolescence: influence of school context. *Journal of Clinical Child & Adolescent Psychology*, (ahead-of-print), 1-15.

Brener, N., Billy, J., & Grady, W. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: evidence from the scientific literature. *Journal of Adolescent Health*, 33, 436-457.

Brown, J., Kranzler, H. R., & Boca, F. K. (1992). Self-reports by alcohol and drug abuse inpatients: factors affecting reliability and validity. *British Journal of Addiction*, 87, 1013-1024.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81.

Centers for Disease Control and Prevention (CDC). (2009). School connectedness: Strategies for increasing protective factors among youth. *Atlanta, GA: US Department of Health and Human Services*.

Centers for Disease Control and Prevention. (2010). Youth Risk Behavior Surveillance — United States, 2009. Surveillance Summaries. *Morbidity and Mortality Weekly Report*, *59,* 1–142.

Chan, D. (2009). So why ask me? Are self-report data really that bad? (2009). In C. E. Lance, & R. J. Vandenberg (Eds.). *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*, 309-336, Taylor & Francis.

Christenson, S., Reschly, A. L., & Wylie, C. (2012). *Handbook of research on student engagement*. New York, NY: Springer.

Cohen, A. C. (1967). Estimation in mixtures of two normal distributions. *Technometrics*, 9 (1), 15-28.

Cohen, J. (2014). School Climate Policy and Practice Trends: A Paradox. A Commentary . *Teaceers College Record*, Date Published: February 21, 2014 http://www.tcrecord.org

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

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* (Vol. 718). John Wiley & Sons.

Cook, C. R., Williams, K. R., Guerra, N. G., Kim, T. E., & Sadek, S. (2010). Predictors of bullying and victimization in childhood and adolescence: A meta-analytic investigation. *School Psychology Quarterly*, *25*, 65.

Cornell, D., Huang, F., Konold, T., Meyer, P., Lacey, A., Nekvasil, E., Heilbrun, A., & Shukla, K. (2013). *Technical Report of the Virginia Secondary School Climate Survey: 2013 Results for 7th and 8th Grade Students and Teachers. Executive Summary.* Charlottesville, VA: Curry School of Education, University of Virginia.

Cornell, D., Huang, F., Konold, T., Meyer, P., Shukla, K., Heilbrun, A., Datta, P., Berg, J., Lacey, A., & Nekvasil, E. (2014). *Technical Report of the Virginia Secondary School Climate Survey: 2014 Results for 9th - 12th Grade Students and Teachers.* Charlottesville, VA: Curry School of Education, University of Virginia.

Cornell, D., Klein, J., Konold, T., & Huang, F. (2012). Effects of validity screening items on adolescent survey data. *Psychological Assessment*, 24, 21.

Cornell, D. G., & Loper, A. B. (1998). Assessment of violence and other high-risk behaviors with a school survey. *School Psychology Review*.

Cornell, D. G., Lovegrove, P. J., & Baly, M. W. (2013). Invalid survey response patterns among middle school students. *Psychological Assessment*. 26, 277-287. doi: 10.1037/a0034808

Cornell, D., & Mayer, M. (2010). Why do school order and safety matter? *Educational Researcher*, *39*, 7-15.

Cornell, D., **Shukla, K**., & Konold, T. (2015). Peer Victimization and Authoritative School Climate: A Multilevel Approach. *Journal of Educational Psychology*. Advance online publication. http://dx.doi.org/10.1037/edu0000038

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*, 98.

Cote, J. A., & Buckley, M. R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research*, 315-318.

Craig, W., Harel-Fisch, Y., Fogel-Grinvald, H., Dostaler, S., Hetland, J., Simons-Morton, B., ... & Pickett, W. (2009). A cross-national profile of bullying and victimization among adolescents in 40 countries. *International Journal of Public Health*, *54*, 216-224.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. John Wiley & Sons.

Cross, J. E., & Newman-Gonchar, R. (2004). Data quality in student risk behavior surveys and administrator training. *Journal of School Violence*, 3, 89-108.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of consulting psychology*, 24, 349.

Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika*, 56, 463-474.

Dayton, C. M. (1991). Educational Applications of Latent Class Analysis. *Measurement and Evaluation in Counseling and Development*, *24*, 131-41.

Dedrick, R.F., & Greenbaum, P.E. (2011). Multilevel confirmatory factor analysis of a scale measuring interagency collaboration of children's mental health agencies. *Journal of Emotional and Behavioral Disorders*, *19*, 27-40. DOI: 10.1177/1063426610365879

Del Boca, F. K., & Darkes, J. (2003). The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction*, *98*, 1-12

Del Boca, F. K., & Noll, J. A. (2000). Truth or consequences: the validity of self-report data in health services research on addictions. *Addiction*, *95*, 347-360.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal statistical Society*, *39*, 1-38.

Diebolt, J., & Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, 363-375.

Dynarski, M., Clarke, L., Cobb, B., Finn, J., Rumberger, R., & Smink, J. (2008). Dropout Prevention. IES Practice Guide. NCEE 2008-4025. *National Center for Education Evaluation and Regional Assistance*.

Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Ross, J., Hawkins, J.,... Wechsler, H. (2008). *Youth Risk Behavior Surveillance — United States, 2007*. MMWR Surveillance Summaries, *57*, 1–131.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Esposito, C. (1999). Learning in urban blight: School climate and its effect on the school performance of urban, minority, low-income children. *School Psychology Review*.

Everitt, B. S. (1996). An introduction to finite mixture distributions. *Statistical Methods in Medical Research*, *5*, 107-127.

Fan, X., Miller, B. C., Park, K. E., Winward, B. W., Christensen, M., Grotevant, H. D., & Tai, R. H. (2006). An exploratory study about inaccuracy and invalidity in adolescent self-report surveys. *Field Methods*, *18*, 223-244.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, *6*, 56-83.

Gerlinger, J., & Wo, J. (2014). Preventing School Bullying: Should Schools Prioritize an Authoritative School Discipline Approach Over Security Measures?. *Journal of School Violence*, (just-accepted).

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, 389-409.

Gladden, R. M., Vivolo-Kantor, A. M., Hamburger, M. E., & Lumpkin, C. D. (2014). *Bullying surveillance among youths: Uniform definitions for public health and recommended data elements, Version 1.0*. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention and US Department of Education.

Goodenow, C. (1993). Classroom belonging among early adolescent students relationships to motivation and achievement. *The Journal of Early Adolescence*, *13*, 21-43.

Goodman, L.A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: A modified latent structure approach. *American Journal of Sociology*, *79*, 1179-1259.

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*, 412-444.

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

Gregory, A., Cornell, D., & Fan, X. (2012). Teacher safety and authoritative school climate in high schools. *American Journal of Education, 118*, 401-425.

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*, 483-496. doi: 10.1037/a0018562

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*, 255-282.

Hansen, T. B., Steenberg, L. M., Palic, S., & Elklit, A. (2012). A review of psychological factors related to bullying victimization in schools. *Aggression and violent behavior*, *17*, 383-387.

Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. *Technometrics*, 8, 431-444.

Hasselblad, V. (1969). Estimation of finite mixtures of distributions from the exponential family. *Journal of the American Statistical Association*, *64*, 1459-1471.

Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, 64, 144.

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.

Janosz, M., Archambault, I., Morizot, J., & Pagani, L. S. (2008). School engagement trajectories and their differential predictive relations to dropout. *Journal of Social Issues*, *64*, 21-40.

Johnson, S. L. (2009). Improving the school environment to reduce school violence: A review of the literature. *Journal of School Health*, *79*, 451-465.

Juvonen, J., Nishina, A., & Graham, S. (2001). Self-views versus peer perceptions of victim status among early adolescents. In J. Juvonen & S. Graham (Eds.), *Peer harassment in school: A plight of the vulnerable and victimized* (pp. 105-124). New York: Guilford Press.

Kendler, K. S., Eaves, L. J., Walters, E. E., Neale, M. C., Heath, A. C., & Kessler, R. C. (1996). The identification and validation of distinct depressive syndromes in a population-based sample of female twins. *Archives of general psychiatry*, *53*, 391.

King, M. F., & Bruner, G. C. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology & Marketing*, *17*, 79-103.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford press.

Klein, J., Cornell, D., & Konold, T. (2012). Relationships between school climate and student risk behaviors. *School Psychology Quarterly, 27*, 154-169.

Konold, T.R., & Cornell, D. (under review). Measurement and structural relations of an authoritative school climate model: A multi-level latent variable investigation. Journal of School Psychology.

Konold, T., Cornell, D., Huang, F., Meyer, P., Lacey, A., Nekvasil, E., Heilbrun, A., & Shukla, K. (2014). Multi-level multi-informant structure of the Authoritative School Climate Survey. *School Psychology Quarterly*.

Konold, T. R., & Glutting, J. J. (2008). ADHD and Method Variance A Latent Variable Approach Applied to a Nationally Representative Sample of College Freshmen. *Journal of Learning Disabilities, 41*, 405-416.

Konold, T. R., & Shukla, K. D. (2014). Informant effects on behavioral and academic associations: A latent variable longitudinal examination. *Psychology in the Schools, 51, 554-566*.

Kuperminc, G. P., Leadbeater, B. J., & Blatt, S. J. (2001). School social climate and individual differences in vulnerability to psychopathology among middle school students. *Journal of School Psychology*, *39*, 141-159.

Lalkhen, A. G., & McCluskey, A. (2008). Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia, Critical Care & Pain*, *8*, 221-223.

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*, 1049-1065.

Lanza, S. T., Coffman, D. L., & Xu, S. (2013). Causal inference in latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*, 361-383.

Lanza, S. T., Dziak, J. J., Huang, L., Wagner, A. T., & Collins, L. M. (2013). *Proc LCA & Proc LTA users' guide (Version 1.3.0).* University Park: The Methodology Center, Penn State. Available from methodology.psu.edu

Lauritsen, J., & Swicegood, C. (1997). The consistency of self-reported initiation of sexual activity. *Family Planning Perspectives*, *29*, 215-221.

Larzelere, R. E., Morris, A. S., & Harrist, A. W. (2013). *Authoritative parenting: Synthesizing nurturance and discipline for optimal child development*. Washington, DC: American Psychological Association.

Leadbeater, B. J., Thompson, K., & Sukhawathanakul, P. (2014). It gets better or does it? Peer victimization and internalizing problems in the transition to young adulthood. *Development and psychopathology*, *26*, 675-688.

Lee, J. S. (2012). The effects of the teacher–student relationship and academic press on student engagement and academic performance. *International Journal of Educational Research*, *53*, 330-340.

Libbey, H. P. (2004). Measuring student relationships to school: Attachment, bonding, connectedness, and engagement. *Journal of School Health*, *74*(7), 274-283.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*, 767-778.

Losen, D.J. (2011). *Discipline Policies, Successful Schools, and Racial Justice*. Boulder, CO: National Education Policy Center. Retrieved [date] from http://nepc.colorado.edu/publication/discipline-policies.

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological methods*, *10*, 21.

Mayworm, A., Sharkey, J., & Parker, R. (2013). School discipline styles and student perceptions of school safety: A latent class analysis. *CASP Today*, *63*, 6-9. Retrieved from: http://www.casponline.org/pdfs/pdfs/ct-fall2013_web.pdf

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics*, 318-324.

McLachlan, G., & Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

Mehta, S. B., Cornell, D., Fan, X., & Gregory, A. (2013). Bullying Climate and School Engagement in Ninth-Grade Students. *Journal of school health*, *83*, 45-52.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543.

Meyer, J. P. (2010). A mixture Rasch model with item response time components. *Applied Psychological Measurement*, *37*, 521-538.

Meyer, P. (2010). *Understanding measurement: reliability*. New York: Oxford University Press, Inc.

Muthén, B., & Asparouhov, T. (2009). Multilevel regression mixture analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172, 639-657.

Muthén, L. K., & Muthén, B. O. (1998–2014). *Mplus user's guide (7th ed.).* Los Angeles, CA: Muthén & Muthén.

Nakamoto, J., & Schwartz, D. (2010). Is Peer Victimization Associated with Academic Achievement? A Meta-analytic Review. *Social Development*, *19*, 221-242.

Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment. *Jama*, *285*, 2094-2100.

National Center for Education Statistics (2005). *School crime supplement to the national crime victimization survey 2005*. Retrieved from http://nces.ed.gov/Programs/Crime/surveys.asp

National School Climate Council (2007). The School Climate Challenge: Narrowing the Gap Between School Climate Research and School Climate Policy, Practice Guidelines and Teacher Education Policy. Available online at: **http://nscc.csee.net/ or http://www.ecs.org/school-climate**

Nylund, K. L. (2007). *Latent transition analysis: Modeling extensions and an application to peer victimization* (Doctoral dissertation, UNIVERSITY OF CALIFORNIA Los Angeles).

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14,* 535-569.

Nylund, K., Bellmore, A., Nishina, A., & Graham, S. (2007). Subtypes, severity, and structural stability of peer victimization: what does latent class analysis say? *Child Development*, *78,* 1706-1722.

Olweus, D. (2007). *The Olweus Bullying Questionnaire*. Center City, MN: Hazelden.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 71-110.

Pellegrini, A. D., Bartini, M., & Brooks, F. (1999). School bullies, victims, and aggressive victims: Factors relating to group affiliation and victimization in early adolescence. *Journal of Educational Psychology*, *91*, 216 –224. doi:10.1037/0022-0663 .91.2.216

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.

Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology*, *78*, 582.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, *12*, 531-544.

Radhakrishna Rao, C. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society-Series B: Statistical Methodology*, *10*, 159-203.

Radliff, K. M., Wheaton, J. E., Robinson, K., & Morris, J. (2012). Illuminating the relationship between bullying and substance use among middle and high school youth. *Addictive Behaviors.* doi: 10.1016/j.addbeh.2012.01.001

Raju, N. S., Price, L. R., Oshima, T. C., & Nering, M. L. (2007). Standardized conditional SEM: A case for conditional reliability. *Applied Psychological Measurement*, *31*, 169-180.

Raudenbush, S. W., Rowan, B., & Kang, S. J. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to US high-school data. *Journal of Educational and Behavioral Statistics*, *16*, 295-330.

Reijntjes, A., Kamphuis, J. H., Prinzie, P., Boelen, P. A., Van der Schoot, M., & Telch, M. J. (2011). Prospective linkages between peer victimization and externalizing problems in children: a meta-analysis. *Aggressive behavior*, *37*, 215-222.

Reijntjes, A., Kamphuis, J. H., Prinzie, P., & Telch, M. J. (2010). Peer victimization and internalizing problems in children: A meta-analysis of longitudinal studies. *Child abuse & neglect*, *34*, 244-252.

Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., ... & Udry, J. R. (1997). Protecting adolescents from harm: findings from the National Longitudinal Study on Adolescent Health. *Jama*, *278*, 823-832.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*.

Rindskopf, D., & Rindskopf, W. (1986). The value of latent class analysis in medical diagnosis. *Statistics in Medicine*, *5*, 21-27.

Robers, S., Kemp, J., and Truman, J. (2013). *Indicators of School Crime and Safety: 2012* (NCES 2013-036/NCJ 241446). National Center for Education Statistics, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC. Retrieved from (August 31, 2014): http://files.eric.ed.gov/fulltext/ED543705.pdf

Robinson-Cimpian, J. P. (2014). Inaccurate Estimation of Disparities Due to Mischievous Responders Several Suggestions to Assess Conclusions. *Educational Researcher*, *43*, 171-185.

Rosenblatt, J., Furlong, M. (1997). Assessing the reliability and validity of student self-reports of Campus violence. *Journal of Youth and Adolescence*, *26*, 187-202.

Sackett, P. R., & Larson Jr, J. R. (1990). Research strategies and tactics in industrial and organizational psychology.

Sharpley, C. F., & Christie, D. R. (2007). 'How I was then and how I am now': current and retrospective self-reports of anxiety and depression in Australian women with breast cancer. *Psycho-oncology*, *16*, 752-762.

Steffgen, G., Recchia, S., & Viechtbauer, W. (2013). The link between school climate and violence in school: A meta-analytic review. *Aggression and Violent Behavior*, *18*, 300-309.

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*, 1266-1281.

Stewart, E. A. (2003). School social bonds, school climate, and school misbehavior: A multilevel analysis. *Justice Quarterly*, *20*, 575-604.

Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, *24*, 236-243.

Stornes, T., Bru, E., & Idsoe, T. (2008). Classroom social structure and motivational climates: On the influence of teachers' involvement, teachers' autonomy support and regulation in relation to motivational climates in school classrooms. *Scandinavian Journal of Educational Research*, *52*, 315-329.

Strömgren, B. (1934). Tables and diagrams for dissecting a frequency curve into components by the half-invariant method. *Scandinavian Actuarial Journal*, *1934*, 7-54.

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

The Pennsylvania State University. (2012). *Definitions*. The Methodology Center. Retrieved from (March 21, 2014): http://methodology.psu.edu/ra/definitions#LCA

Tittle, C. R., & Hill, R. J. (1967). The accuracy of self-reported data and prediction of political activity. Public Opinion Quarterly, 103-106.

Tomlinson, C. A. (1995). Differentiating Instruction for Advanced Learners in the Mixed-Ability Middle School Classroom. ERIC Digest E536.

Van Horn, M. L. (2003). Assessing the unit of measurement for school climate through psychometric and outcome analyses of the school climate survey. *Educational and Psychological Measurement*, *63*, 1002-1019.

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology*, 33, 213–239.

Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, 17, 33–51.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. *Applied Latent Class Analysis*, 89-106.

Vermunt, J. K., & Magidson, J. (2004). Latent class analysis. *The sage encyclopedia of social sciences research methods*, 549-553.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307-333.

Wang, M. T., & Holcombe, R. (2010). Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Research Journal*, *47*, 633-662.

Wang, J., Iannotti, R. J., Luk, J. W., & Nansel, T. R. (2010). Co-occurrence of victimization from five subtypes of bullying: Physical, verbal, social exclusion, spreading rumors, and cyber. *Journal of Pediatric Psychology*, *35*, 1103-1112.

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.

Wedel, M., & DeSarbo, W. S. (1995). A mixture likelihood approach for generalized linear models. *Journal of Classification*, 12, 21-55.

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, 27, 551-572.

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, 18, 163-183.

Wolfe, J. H. (1967). *NORMIX: Computational Methods for Estimation the Parameters of Multivariate Normal Mixtures of Distributions* (No. NPRA-SRM-68-2). Naval Personnel Research Activity, San Diego, CA.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, *5*, 329-350.

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.

Appendix A.

M*plus* Syntax for Study 1: Measurement Application

```
TITLE: 2-Class multilevel LC Model on response inconsistency variables


DATA: FILE = MeasureApp_March2.dat;

VARIABLE: NAMES =   used truthyn   fastresp
  r1 - r7   Gender   white   black   hispanic   asian
  multirace   amrind   parented   grades   acexp
  vict   bullied   bullyoth
  weapon   phyfight   suicidatt   alcohol  marijuana;

        USEVARIABLES = r1 - r7;

! NOTE: r1 to r7 are response-inconsistency variables respectively for-
! structure, academic press, respect for students, willingness to seek help,
! academic engagement, cognitive engagement, prevalence of teasing and bullying (PTB)

  Auxiliary = (e) truthyn   fastresp Gender white black hispanic asian
  multirace amrind parented grades   acexp vict bullied bullyoth
  weapon phyfight  suicidatt alcohol marijuana;

        Missing = *    all(-999);
        CLUSTER= usid;
        CLASSES= c(2);

ANALYSIS:
TYPE = TWOLEVEL MIXTURE ; ! Multilevel latent class analysis
STARTS = 200 50; ! 200 random starts and 50 optimizations help obtain stable estimates
LRTBOOTSTRAP =300;
Process = 2;


OUTPUT:  TECH11 TECH14;

PLOT: TYPE = Plot3;
      SERIES= r1 -r7 (*);

SAVEDATA:
    FILE = study1_out2c.dat;
    FORMAT = free;
    SAVE = CPROB;
```

Appendix B.

Mplus syntax for Study 2: Substantive Application

```
TITLE: 4-Class multilevel LC modelling on school climate measures

DATA: FILE = SubstantiveApp_March3.dat;

VARIABLE:

NAMES = usid   gender   ped

expect grades ! Academic outcomes

white   black   hisp   asian   multirac   amrind  ! race/ethnicity dummy variables

affeng   cogeng   struc   acapres   rspt   wsh
ptb   bullyvic   vict   ! Eight continuous school climate measures

bullyvic   bullyoth ! Two dichotomous school climate measure of bullying experience

weapon   phyfight   suicide   alcohol   mariju; !Five risk behaviors

Missing = All (-999);

USEVARIABLES = affeng   cogeng   struc   acapres   rspt
  wsh   ptb  vict   bullyvic  bullyoth ;

AUXILIARY = (e) gender ped expect grades
black white hisp asian multirac amrind
weapon   phyfight   suicide   alcohol   mariju;

Categorical= bullyvic   bullyoth;
Cluster = usid;
WITHIN= affeng cogeng struc  acpres
            rspt wsh ptb vict bullyvic bullyoth ;
CLASSES= c(4);

  ANALYSIS: TYPE = TWOLEVEL MIXTURE;
        STARTS = 200 50;
        LRTBOOTSTRAP =300;
        Process = 2;

Model:
    %WITHIN%
     %OVERALL%
```

```
    %C#1%
affeng cogeng struc  acpres
            rspt wsh ptb vict;
        %C#2%
affeng cogeng struc  acpres
            rspt wsh ptb vict;

        %C#3%
affeng cogeng struc  acpres
            rspt wsh ptb vict;

        %C#4%
affeng cogeng struc  acpres
            rspt wsh ptb vict;

OUTPUT:  TECH11 TECH14;

PLOT: TYPE = Plot3;
       SERIES=   affeng
  cogeng   struc  acapres   rspt
  wsh   ptb  vict(*);

SAVEDATA:
    FILE = study2_mar3_out.dat;
    FORMAT = free;
    SAVE = CPROB;
```