Article

Educational and Psychological Measurement
2020, Vol. 80(4) 617–637

DOI: 10.1177/0013164419885999
journals.sagepub.com/home/epm
SAGE

Using Integrative Data Analysis to Investigate School Climate Across Multiple Informants

Kathleen V. McGrath[1], Elizabeth A. Leighton[1], Mihaela Ene[1], Christine DiStefano[1] and Diane M. Monrad[1]

Abstract

Survey research frequently involves the collection of data from multiple informants. Results, however, are usually analyzed by informant group, potentially ignoring important relationships across groups. When the same construct(s) are measured, integrative data analysis (IDA) allows pooling of data from multiple sources into one data set to examine information from multiple perspectives within the same analysis. Here, the IDA procedure is demonstrated via the examination of pooled data from student and teacher school climate surveys. This study contributes to the sparse literature regarding IDA applications in the social sciences, specifically in education. It also lays the groundwork for future educational researchers interested in the practical applications of the IDA framework to empirical data sets with complex model structures.

Keywords

integrative data analysis, survey research, school climate

Surveys are used extensively in the social sciences to collect data from multiple informants and provide information about target areas or constructs. Inherent differences across informant groups often necessitate different items to reflect unique perceptions of a group. For instance, a teacher survey about school climate may contain items regarding working conditions and students may receive items addressing course rigor.

[1]University of South Carolina, Columbia, SC, USA

Corresponding Author:
Kathleen V. McGrath, University of South Carolina, Wardlaw Building, Suite 014, 820 Main Street, Columbia, SC 29208, USA.
Email: mcgrathv@email.sc.edu

Differences in survey content help researchers to gain a deeper understanding of how a group perceives the construct under study and facilitates a holistic view of the construct itself.

From a measurement perspective, analyzing data separately when data were collected from multiple informants can present methodological challenges or limitations that may paint an incomplete, or even biased, picture of a construct (Bainter & Curran, 2015). One technique often used by researchers to disentangle the effects of similar constructs measured across different respondent groups is the confirmatory factor analysis (CFA) approach coupled with a multitrait–multimethod model perspective. However, multitrait–multimethod models typically require at least three different methods measuring the same three traits for model identification (Hox & Bechger, 1995). Additionally, longitudinal models, such as latent growth curve models, assume that the same individuals are measured at each time point using the same metric (Widaman, Ferrer, & Conger, 2010). Survey findings may also be limited by the sample characteristics, such as underrepresentation of important subgroups (e.g., ethnic minorities).

A potential solution for these challenges in survey research lies under the integrative data analysis (IDA) framework which permits the pooling of multiple surveys under one model (Bauer & Hussong, 2009). This study is unique in that it presents an application of the IDA framework within the context of pooled student and teacher school climate survey data. Furthermore, it is one of the first in the education field to employ a practical application of the IDA framework to an empirical data set with multiple factors.

Integrative Data Analysis

IDA was developed out of the meta- and mega-analysis literature as a way to pool raw data across multiple, independent studies and fit a model directly to the combined data set (Bauer & Hussong, 2009; Curran et al., 2008; Curran et al., 2018; Curran & Hussong, 2009). Meta-analysis involves the pooling of results/parameter estimates across studies to assess the cumulative findings (Curran et al., 2008; Glass, 1976). Mega-analysis procedures expand on this concept by allowing for the simultaneous analysis of raw data from multiple studies (see Carlson & Miller, 1987; Cialdini & Fultz, 1990; McArdle & Horn, 2002).

At the root of meta- and mega-analytic procedures is data fusion, which emerged in the early 1960s and 1970s (Rässler, 2003). Data fusion is a process wherein data from multiple sources with common variables or constructs are combined with the intent to create a single, complete data source that contains information on all variables (Marcoulides & Grimm, 2017; Rässler, 2003). More recent data fusion research has centered on the application of latent variable models to combined data sets (e.g., Curran et al., 2008; Hussong, Flora, Curran, Chassin, & Zucker, 2008; Marcoulides & Grimm, 2017; McArdle, Grimm, Hamagami, Bowles, & Meredith, 2009). Measurement models within the structural equation modeling and item response theory frameworks as well as factor analytic models such as moderated nonlinear factor analysis (MNLFA) can be used (Bainter & Curran, 2015).

The methods adopted in the current article can be considered as data fusion techniques combined with the application of an MNLFA model. While these individual methodological components exist, we hold that they do so within an analytical framework called integrative data analysis. The IDA framework should not be considered as a singular ''standardized technique'' but as a guide for researchers to successfully conduct secondary data analysis (Bainter & Curran, 2015, p. 2). This framework appears often in the literature (Bainter & Curran, 2015; Bauer & Hussong, 2009; Curran et al., 2008; Curran et al., 2014; Curran et al., 2018; Curran & Hussong, 2009; Hofer & Piccinin, 2009). The current study was designed to apply the framework to an empirical data set and a ''real-life'' example that was atypical to illustrations of the framework as presented by Curran et al. (2014) and Bauer (2017). For this reason, we acknowledge the adopted terminology in these works and apply the term ''IDA'' in an analogous context throughout the article.

The general purpose of IDA is to fit a statistical model to data that have been combined from separate data samples or sources, providing a broader and more comprehensive view of the latent construct (Curran & Hussong, 2009). Under this framework, data obtained from different sources and/or using different measurement tools can be combined as long as the purpose is to measure the same overarching construct (Bauer & Hussong, 2009; Curran & Hussong, 2009). While identical measurement tools are not mandatory, IDA requires that participants across studies or groups be measured with a subset of common items that share both theoretical and operational definitions of the same construct (Bauer & Hussong, 2009). Within the IDA framework, scale scores utilizing data across sources can be created, and issues of measurement invariance between sources can be investigated.

A central step within our adopted IDA procedure is the use of MNLFA. This method is used to examine and identify issues of measurement invariance as it relates to covariates of interest (e.g., study or group membership), and to create commensurate scale scores that account for all items (both common and unique) in the pooled data (Bauer & Hussong, 2009). With MNLFA, the model can simultaneously take into account the potential for multiple response options across measures used in different studies or with different groups, as well as the influence of covariates on factor means/variances and item intercepts/loadings (Curran et al., 2014).

It is also the MNLFA method that allows for the evaluation of measurement invariance which is a precondition for group comparison. Measurement invariance exists when values for item responses rely solely on the values of the latent variable and, in turn, the item should have the same meaning across groups (Bauer, 2017). Varied scores across groups suggest that the latent construct holds different meanings for different groups of respondents. For example, in the case of school climate, any differences in item responses should reflect only respondents' perceptions of school climate factors (e.g., home–school relations, physical environment), and not group membership. The IDA framework accommodates the detection of measurement invariance and facilitates group comparison, as it assumes shared latent variables between respondent groups.

Curran and Hussong (2009) expanded on the benefits of IDA, citing its ability to assess the ''equivalence at multiple levels of design that can often incorporate differences in sampling, geographic region, history, assessment protocol, psychometric measurement, and even hypothesis testing'' (p. 86). Furthermore, the ability to combine multiple data sets not only increases statistical power but also can add greatly to sample heterogeneity, incorporating important subgroups based on factors such as gender or socioeconomic status. As a result of pooling data, researchers can also increase the observed number of low-frequency behaviors, increasing the stability of model estimation (Curran & Hussong, 2009).

IDA can also broaden the operationalization of constructs. While a series of studies may define constructs differently, combining data from these studies enhances construct validity and fosters the increased generalizability of results. Furthermore, IDA facilitates the sharing of data and increases efficiency, as researchers can collaborate to accumulate data and further study of an area (Curran & Hussong, 2009).

Finally, IDA allows for increased sample diversity and representation of minority subgroups, the representation of multiple measurement techniques to curtail weaknesses of any one study, and the ability to measure constructs and how they change across studies and time (Bainter & Curran, 2015). These attributes widen the scope of available data from previous high-quality studies for researchers, making data more accessible, affordable, and transparent.

Implications of IDA for Survey Research

Researchers often develop and use surveys to measure an underlying construct of interest. To attain a comprehensive view of this construct, multiple stakeholder groups are often assessed. Unfortunately, while surveys are generally cost-effective measures, the studies which utilize them (e.g., therapeutic or educational interventions) can be costly, time-intensive, and reliant on unpredictable external funding—all of which can hinder the replication of studies that would otherwise support the generalizability of previous findings (Bauer & Hussong, 2009). Low response rates can also be problematic. For example, with school surveys, parent response rates can be extremely low and vary widely across grade levels and each school's delivery and collection methods. Low response rates hinder the generalizability of findings and prevent schools from tracking progress over time.

Integrative data analysis would allow for the compilation of data across various timepoints, informants, and instruments in survey research, alleviating many of the difficulties encountered with survey research. Researchers could incorporate data from previous studies into their work, expanding on the resources available to them as well as their potential findings. Although a possible limitation may be the high interstudy heterogeneity of methodology, collaboration and cooperation among researchers to improve study integration techniques could assuage possible difficulties that may arise from differences in design, measurement, sampling techniques, and the like (Curran & Hussong, 2009). While IDA has many benefits, it has not been utilized much in the field of education. The current study presents an application of the IDA framework within the context of pooled student and teacher school climate survey data.

The Context of School Climate

Surveys have become prevalent in the measurement of student, teacher, and parent perceptions of school climate (Wang & Degol, 2016). Not only are the benefits of positive school climate well-documented, but they have been a focal point of educational policy and reform for the past three decades (Thapa, Cohen, Guffey, & Higgins-D'Alessandro, 2013). Favorable school climate is associated with improved student academic achievement (Greenberg, 2004; Lee & Burkham, 1996; Stewart, 2007), increased teacher job satisfaction (Ma & MacMillan, 1999), and enhanced home–school relationships (DiStefano, Monrad, May, McGuiness, & Dickenson, 2007). Furthermore, the manageable nature of school climate compared with other barriers to academic achievement (e.g., school poverty levels) makes it a popular target of school reform initiatives among policy makers and researchers.

In the state of South Carolina (SC), school climate surveys are administered in all public schools which then use results to gauge yearly progress, inform future decisions, and meet requirements of the state's accountability legislation. At present, SC school climate surveys have been analyzed separately by group (i.e., students, teachers, parents), including studies identifying the factor structure within each group and studies of measurement invariance within teachers (DiStefano, Mindrila, & Monrad, 2013; Monrad et al., 2008). However, survey data have never been pooled across informant groups.

One example of how the IDA framework can be applied with this example concerns meeting federal accountability standards outlined by the Every Student Succeeds Act (ESSA). Signed into law in 2015, replaces No Child Left Behind (NCLB) and gives states more freedom in selecting accountability goals and systems. However, states are required to include a minimum of one nonacademic indicator of school quality to measure improvements in areas such as school equity and climate (Penuel, Meyer, & Valladares, 2016).

Given the linkages between a positive school climate and improved student and teacher outcomes, it is essential that current and future research equips stakeholders to make advancements in the field. The IDA framework expands the breadth and depth of a chosen construct via increased sample heterogeneity and size, improved capacity to integrate and replicate findings, and a better understanding of how constructs evolve over time (Bainter & Curran, 2015). It is this expansion that makes IDA a unique and feasible framework to explore in an applied setting.

The purpose of this study was to present an application of the IDA using a sample of student and teacher school climate surveys. Applying this framework, the study sought to (a) investigate measurement invariance across groups, (b) create and analyze scale scores using combined teacher and student data, and (c) examine the relationships between these scores and relevant outcome variables (e.g., test scores). While the current study applied the IDA framework to school climate, the practical applications of IDA extend far beyond a single construct and may provide a more holistic view of school climate.

Method

Data Sources

This study utilized school climate data collected by the SC State Department of Education in the spring of 2016. While the state of South Carolina collects school climate data annually from teachers, students, and parents, only data from students and teachers were used due to high levels of parent nonresponse. Students and their parents are selected from the highest grade levels at each school (typically 5, 8, and 11 for elementary, middle, and high schools, respectively), while teachers across all grade levels within a school are targeted. The teacher survey contains 81 questions about the school's learning environment, home–school relationships, social–physical environment, and working conditions. The student survey contains 51 questions that address learning environment, social–physical environment, and home–school relations.

Currently, South Carolina includes measures of school climate on each school's report card for accountability. However, the information included on the report cards only includes the percentage agreement across three survey items from parents, teachers, and students regarding their satisfaction with the learning environment, the social and physical environment, and school–home relations.

The current study used teacher and student climate data collected from 199 high schools (31,853 students and 7,884 teachers) across South Carolina. The population of interest was considered to be all high school teachers and students in South Carolina. The student and teacher surveys consist of Likert-type scale items (1 = *disagree*, 2 = *mostly disagree*, 3 = *mostly agree*, 4 = *agree*). Previous exploratory factor analysis (EFA) and CFA studies of the state's school climate surveys have identified and replicated a stable factor structure over time, supporting the application of factor scores in further analyses (DiStefano et al., 2007; Monrad et al., 2008).

Data Analysis

IDA and MNLFA Procedures. A primary challenge with IDA is the estimation of valid and reliable factor scores that originate from data which may consist of different items, contain varying response options, and originate from multiple studies (Curran et al., 2014). Curran et al. (2014) developed a general framework for IDA measurement designs and obtaining these scores. The proposed framework consists of five steps: (a) conduct graphical and descriptive analyses of individual items, (b) test dimensionality through nonlinear exploratory factor analysis, (c) test for factor and item differences via MNLFA, (d) estimate scale scores in an integrated data set, and (e) evaluate the quality of the final scale scores.

The MNLFA procedure as introduced by Bauer (2017) was used to test for factor and item differences (noted above in Step c). Stages within this procedure include (a) the determination of factor structure; (b) the fitting of MNLFA models to each factor, including mean and variance specification and differential item functioning (DIF) detection; and (c) the fitting of a multidimensional MNLFA model. All analyses were conducted using *Mplus* v8.0 (Mutheän & Mutheän, 2017).

MNLFA combines elements of both multiple group and multiple-indicator multicause (MIMIC) modeling procedures. Like the multiple group model, MNLFA allows for the moderation of variance and covariance parameters in addition to the means and intercepts which are typically not included in a MIMIC model. Nonuniform DIF is also permitted through the moderation of factor loadings. Similar to MIMIC models, MNLFA can evaluate model invariance and DIF as a function of multiple covariates which can be either continuous or categorical in nature (Bauer, 2017). Together, these characteristics permit the location of uniform and nonuniform DIF.

Missing data were considered to be missing at random, given that missingness in the climate results was the result of the group covariate (i.e., teacher or student; Muthén & Muthén, 1998-2012). While data were present for shared items, data were missing for student observations on unique teacher items and for teacher observations on unique student items. This pattern of missingness is common in IDA given the pooling of items across different instruments, and the data for unique items are referred to as ''missing by design'' (Bauer & Hussong, 2009, p. 106).

While WLSMV (mean- and variance-adjusted weighted least squares estimator) is the preferred estimator with ordinal data (Curran et al., 2014); this choice prompts the pairwise deletion of missing data in *Mplus* (Asparouhov & Muthén, 2010). Although Likert-type data are used with the climate surveys, the structure of missing data in this study did not permit the use of pairwise deletion because data for group-specific items would have been removed. Instead, the maximum likelihood estimator was used for all procedures, as this method uses full-information maximum likelihood (FIML) in the presence of missing data. This decision was further supported by research showing that the presence of four or more item response categories yields similar results if data are symmetric, regardless of whether categorial or continuous estimation procedures are used (Finney & DiStefano, 2013).

Preliminary Data Preparation. Prior to running the initial analysis, we conducted a content analysis of the survey items to identify the common and unique items included in the survey. This procedure identified 27 common items shared between the teacher and student climate surveys as well as 20 unique items from the student survey and 48 unique items from the teacher survey. Of the 27 common items, 13 were identical items and 14 items lacked equivalent wording but were comparable in content. For example, concerning the physical condition of a school, the student item ‘‘Broken things at my school get fixed’’ would be linked to the teacher item ‘‘The school building is maintained well and repaired when needed.’’

Results

Results are provided sequentially to facilitate interpretation of steps involved with the IDA process.

Step 1: Descriptive Analysis of Items

A descriptive analysis of items was conducted across the entire data set and by respondent group. Data from each group were labeled (common, teacher, student) and compiled into a single data set. Low- and/or cross-loading items were removed after conducting exploratory factor analysis on the pooled data set. As a result, the total number of items was reduced to 45 items with 16 common items (9 identical, 7 similar), 15 unique items from the student survey, and 14 unique items from the teacher survey. Descriptive statistics for the 16 common items between the teacher and student surveys were analyzed to determine if they behaved similarly across groups; these items are denoted with the prefix ‘‘C’’ (see Table 1). Across items, mean responses were largely comparable between teachers and students, with students exhibiting a slightly lower item average across all items. Student data were more normally distributed, with skewness and kurtosis values within a normal range ($< \pm 3$ and $< \pm 7$, respectively; Kline, 2005; West, Finch, & Curran, 1995). The teacher data were slightly more problematic with some items displaying nonnormality (i.e., kurtosis values exceeding ± 7).

Regarding unique teacher items that were assigned a ‘‘T’’ prefix, descriptive analysis revealed fairly stable item means centered at a value of 3.07 (see Table 2). Mean values were relatively low for Item 1 (0.80), Item 18 (1.6), and Item 19 (1.6). The average standard deviation across teacher items was 0.75, with values fairly comparable in size. Skewness values were all within an acceptable range, with only one item exceeding the lower limit of ± 2 (Item 20 = -2.26). Nonnormality for teacher items was also present in the unique data set for several items (n = 5); however, on average, kurtosis values were relatively symmetric.

Descriptive statistics for the unique student items were also analyzed (see Table 3), and all items were assigned an ‘‘S’’ prefix. Items means were stable and centered on a value of 3.05, similar to that for all unique teacher items. Standard deviation values ranged from 0.68 to 1.10 and were generally larger for unique student items compared with unique teacher items, averaging at about 0.88. Skewness and kurtosis values for the unique student items were all within acceptable range except for Item 8, which had a kurtosis value of 2.5.

Table 1. Descriptive Statistics for Student and Teacher Common Survey Items.

	Student				Teacher			
Item	*M*	*SD*	Skewness	Kurtosis	*M*	*SD*	Skewness	Kurtosis
C1. Clean grounds	2.96	0.97	−0.69	−.46	3.59	0.68	−1.80	3.25
C2. Clean hallways	3.08	0.92	−0.86	−.02	3.61	0.65	−1.89	3.72
C3. Clean bathrooms	2.45	1.07	0.01	−1.25	3.40	0.82	−1.37	1.25
C4. Class behavior	2.52	0.95	−0.20	−.90	3.14	0.81	−.89	0.56
C5. Hall behavior	2.46	0.97	−0.10	−1.01	3.04	0.87	−.78	0.05
C6. Safe before/after school	3.24	0.88	−1.11	0.60	3.68	0.61	−2.23	5.56
C7. Safe during school	3.28	0.86	−1.18	0.80	3.72	0.58	−2.41	6.59
C8. Safe to/from school	3.34	0.82	−1.29	1.27	3.78	0.50	−2.76	9.26
C9. Home-school relations	3.17	0.92	−0.99	0.19	3.07	0.88	−.68	−0.27
C10. Parent activities	3.11	0.94	−0.85	−0.19	3.46	0.68	−1.17	1.24
C11. Parent learning expectations	3.14	0.93	−0.91	–0.03	3.24	0.77	−0.82	0.22
C12. Parent homework	2.54	1.07	−0.08	−1.25	2.98	0.81	−0.54	−0.11
C13. Parent volunteering	2.96	1.00	−0.65	−0.65	2.68	1.03	−0.19	−1.12
C14. Understand instruction	3.08	0.86	−0.79	0.08	3.61	0.58	−1.42	2.03
C15. Learning expectations	3.42	0.72	−1.28	1.69	3.62	0.60	−1.52	2.25
C16. School maintenance	2.76	1.00	−0.39	−0.89	3.41	0.80	−1.34	1.23

Note. *M* = mean; *SD* = standard deviation.

Table 2. Descriptive Statistics for Unique Teacher Items.

Item	*M*	*SD*	Skewness	Kurtosis
T1. Parent school policies	0.80	0.40	−1.12	1.09
T2. Parent instruction support	3.40	0.71	−0.74	0.46
T3. Parent conference attendance	3.12	0.77	−0.69	0.07
T4. Parent discipline support	3.08	0.81	−0.73	0.55
T5. Parent meeting attendance	3.10	0.76	−0.46	−0.48
T6. Parent advisory committee	2.95	0.87	−0.79	−0.16
T7. Standard implementation	3.17	0.86	−1.84	3.77
T8. Sufficient teaching time	3.73	0.49	−1.70	2.87
T9. Assessment to instruction	3.60	0.64	−1.34	1.72
T10. Low achiever needs met	3.56	0.62	−1.34	1.56
T11. Effective special education	3.52	0.67	−1.98	4.10
T12. Gifted needs met	3.68	0.60	−1.79	3.13
T13. Motivated students	3.63	0.64	−0.59	−0.15
T16. Sufficient space	3.34	0.93	−1.47	1.50

Note. *M* = mean; *SD* = standard deviation.

Table 3. Descriptive Statistics for Unique Student Items.

Item	*M*	*SD*	Skewness	Kurtosis
S2. Parent homework help	2.88	1.10	−0.58	−1.01
S3. Parent knowledge of academic performance	3.41	0.79	−1.40	1.63
S4. Parents welcomed	3.37	0.78	−1.27	1.38
S5. Interesting classes	2.60	0.95	−0.22	−0.86
S6. Teacher success: ELA	3.32	0.83	−1.21	0.96
S7. Teacher success: Math	3.08	0.98	−0.84	−0.34
S8. Teacher behavior expectations	3.52	0.68	−1.51	2.50
S9. Helpful homework	2.91	0.91	−0.59	−0.39
S10. Representative tests	3.35	0.78	−1.21	1.19
S11. Teacher learning assistance	3.07	0.83	−0.76	0.19
S12. Teacher praise	2.79	0.94	−0.41	−0.70
S13. Teacher time	2.90	0.90	−0.58	−0.35
S14. Student learning confidence	2.75	0.89	−0.40	−0.52
S15. Useful textbooks	2.50	1.01	−0.07	−1.09
S16. Teacher collaboration	3.00	0.89	−0.70	−0.14

Note. M = mean; *SD* = standard deviation; ELA = English language arts.

Step 2: Determination of Factor Structure

Previous factor analytic work on SC school climate surveys identified four factors in the student survey (learning environment, home–school relationships, social–physical environment, and safety) and six factors in the teacher survey (working conditions and leadership, home–school relationships, instructional focus, resources, physical environment, and safety; see Gareau et al., 2010). Using both factor structures as a starting point, common factors and items were identified across each survey. In considering the dimensionality of our item pool and the correspondence between student and teacher survey items, unique teacher items related to working conditions (with no student parallel) were removed from the analysis.

The factor structure of our data was determined by conducting both EFA and CFA. Given the previous factor analytic work and the consistency of results over many years (see Gareau et al., 2010; Monrad et al., 2008), we did not conduct EFA for the individual teacher and student surveys. Common EFA recommendations for identifying the dimensional structure were used to evaluate alternative models such as factor loadings of .40 or greater, no cross-loadings greater than .30, percentage of variance explained, and interpretability.

Our findings supported the use of a four-factor model which is consistent with previous research concerning the structure of SC school climate surveys (see DiStefano et al., 2007; Ene et al., 2016; Monrad et al., 2008). To substantiate our decision to use a four-factor model, we conducted confirmatory factor analysis using maximum likelihood estimation techniques.

Table 4. Confirmatory Factor Analysis Standardized Loadings for Combined (C), Teacher (T), and Student (S) Items.

Loadings by factor	Estimate	*SE*
Home–school relationships		
C9. Home–school relations	.681	.003
C10. Parent activities	.699	.003
C11. Parent learning expectations	.730	.003
C12. Parent homework	.657	.003
C13. Parent volunteering	.631	.003
T1. Parent school policies	.717	.006
T2. Parent instruction support	.843	.003
T3. Parent conference attendance	.800	.004
T4. Parent discipline support	.819	.004
T5. Parent meeting attendance	.839	.004
T6. Parent advisory committee	.729	.005
T13. Motivated students	.676	.006
S2. Parent homework help	.571	.004
S3. Parent knowledge of academic performance	.676	.003
S4. Parents welcomed	.671	.003
Learning environment		
C14. Understand instruction	.688	.003
C15. Learning expectations	.647	.003
T7. Standard implementation	.770	.005
T8. Sufficient teaching time	.669	.007
T9. Assessment to instruction	.819	.004
10. Low achiever needs met	.830	.004
T11. Effective special education	.680	.007
T12. Gifted needs met	.752	.006
S5. Interesting classes	.588	.004
S6. Teacher success: ELA	.499	.004
S7. Teacher success: Math	.520	.004
S8. Teacher behavior expectations	.502	.004
S9. Helpful homework	.599	.004
S10. Representative tests	.557	.004
S11. Teacher learning assistance	.722	.003
S12. Teacher praise	.607	.004
S13. Teacher time	.721	.003
S14. Student learning confidence	.559	.004
S15. Useful textbooks	.561	.004
S16. Teacher collaboration	.691	.003
Physical environment		
C1. Clean grounds	.815	.002
C2. Clean hallways	.816	.002
C3. Clean bathrooms	.766	.002
C4. Class behavior	.651	.003
C5. Hall behavior	.658	.003
C16. School maintenance	.736	.003
T16. Sufficient space	.665	.008
School safety		
C6. Safe before/after school	.910	.001

(continued)

Table 4. Continued

Loadings by factor	Estimate	*SE*
C7. Safe during school	.935	.001
C8. Safe to/from school	.821	.002

Note. SE = standard error; ELA = English language arts.

Fit statistics for the selected model were comparable to those generated in student and teacher models in past years (comparative fit index [CFI] = .875, root mean square error of approximation [RMSEA] = .053, 90% confidence interval [CI] = [.052, .053], standardized root mean square residual [SRMR] = .074). The cutoff criterion of CFI $\geq$ .95 is typically used to indicate good model fit, with values of CFI $\geq$ .90 indicating that the model has not been misspecified (Hu & Bentler, 1999). Upper limit values for both RMSEA and SRMR are usually set to be less than .06 and .08, respectively, to indicate good model fit (Hu & Bentler, 1999; MacCallum, Browne, & Sugawara, 1996). These values also meet the combination rules of Hu and Bentler's (1999) two-index presentation with an RMSEA value less than .06 and SRMR less than .09. Reliance on a single index to assess model fit is inappropriate; these sentiments are reflected in the frequent reporting of multiple fit indices in academic literature. Considering local fit, standardized loading values from the CFA were all statistically significant and fairly high, ranging from .499 to .930, with values centered on .70. All standardized loadings for the combined CFA model are presented in Table 4.

Step 3: Testing for Factor and Item Differences

Once the dimensional structure was identified, MNLFA was used to examine measurement invariance between teachers and students at the factor and item levels. As specified by Bauer (2017), MNLFA procedures started with the specification of separate MNLFA models for each of the four identified factors. To identify items that would be tested for DIF in the full MNLFA model (all factors combined), items were examined for DIF using an iterative procedure, similar to that in Bauer (2017). Through this procedure, modification indices were investigated to see which items, by allowing DIF, would result in the greatest model fit improvement. First, DIF was modeled for the item with the largest modification index, and then modification indices were reexamined. This process continued item-by-item until no further significant model improvement was identified. Using this procedure, seven items were flagged to be investigated for DIF through MNLFA.

Next, smaller MNLFA models were fit for each factor. These models allowed for covariate effects on factor means and variances, as well as on item intercepts and loadings. Significant covariate effects were found for factor means and variances across item intercepts and loadings on the seven items examined for DIF. Finally, a

Table 5. Parameter Estimates, Final Moderated Nonlinear Factor Analysis Model.

Reference parameter	Covariate effect (*SE*)
Home–school relationships factor	
Mean	−0.640 (0.017)*
Variance	0.144 (0.013)*
Learning environment factor	
Mean	−0.951 (0.020)*
Variance	−0.059 (0.014)*
Physical environment factor	
Mean	−1.024 (0.016)*
Variance	0.239 (0.012)*
Safety factor	
Mean	−0.670 (0.014)*
Variance	0.317 (0.012)*
C2. Clean hallways	
Intercept	0.123 (0.009)*
Loading	0.080 (0.008)*
C3. Clean bathrooms	
Intercept	−0.275 (0.011)*
Loading	−0.002 (0.011)
C8. School to/from	
Intercept	−0.070 (0.006)*
Loading	0.085 (0.007)*
C9. Home–school relations	
Intercept	0.446 (0.011)*
Loading	−0.241 (0.011)*
C11. Parent learning expectations	
Intercept	0.287 (0.010)*
Loading	−0.038 (0.010)*
C13. Parent volunteer	
Intercept	0.652 (0.012)*
Loading	−0.260 (0.012)*
C15. Teacher learning expectations	
Intercept	0.235 (0.010)*
Loading	−0.053 (0.009)*

Note. *SE* = standard error.
*$p < .001$.

full MNLFA model containing all factors and all significant covariate effects identified in the single-factor models was examined.

Results from the full model are found in Table 5 and indicate a statistically significant covariate effect on all factors ($p < .001$), where students exhibited lower perceptions of school climate than teachers. Deviations between student and teacher opinions on school climate were largest for the physical environment factor, with a covariate effect of −1.024. This suggests that students' mean perception of physical environment falls about one standard deviation below the mean of teachers. The covariate effect of group on learning environment was not far behind, with a value of −0.951, followed by the covariate effect on safety and home–school relationships of −0.670 and −0.640, respectively.

Looking to item parameter estimates, all covariate effects on item intercepts and loadings were statistically significant with the exception of the covariate effect on the intercept for Item X3 (clean hallways, p = .882). This suggests the presence of an interaction effect between group, the level of the latent trait, and item responses (Crane, van Belle, & Larson, 2004). Positive covariate effects on item intercepts were also present for most items flagged for DIF. This indicates that for individuals with the same level of the latent variable, students rated these items higher than teachers. Exceptions included items related to the cleanliness of bathrooms and safety to and from school, where teachers rated the items more positively. Covariate effects on the loading values were largest for Item C9 (home–school relations) and Item C13 (parent volunteering) at −0.241 and −0.260, respectively. This suggests that these items have a stronger association to the latent factor for teachers than for students.

Step 4: Factor Scores

Following the identification of factor and item differences through MNLFA, *Mplus* was used to generate factor scores using the maximum a posteriori method (Muthén & Muthén, 2019). A key characteristic of the MNLFA model is that it ''allows for the creation of scale scores based on all available items for a given participant while accounting for potential differences in both the latent factor and the individual items as a function of observed covariates'' (Bauer & Hussong, 2009, as cited in Curran et al., 2014). More specifically, to circumvent possible confounding effects with study group membership, MNLFA simultaneously tests for measurement invariance (i.e., commensurate scaling) across factors by incorporating covariate effects on factor means, factor variances, factor loadings, and item intercepts into the model (Bauer & Hussong, 2009; Curran et al., 2014).

Here, factor scores were created using *Mplus* with the MNLFA model after combining teacher and student data and accounting for covariate effects at the factor and item levels. These scores were then aggregated by school to generate average school-level factor scores. To remain consistent with previous research on the SC school climate surveys, mean factor scores were computed for schools that contained at least 10 teacher responses and 15 student responses (Monrad et al., 2008). This resulted in the creation of factor scores for 180 schools.

The viability of these scores was analyzed by examining their relationship with relevant school outcomes that are highly correlated with teacher and student climate data, including ACT scores, end-of-course (EOC) exam pass rates, graduation rate, and the South Carolina State Poverty Index (see Monrad et al., 2008). To measure the strength of association between the total factor score and academic outcomes, bivariate analyses were conducted on both combined and single-group (teacher, student) data sets to obtain correlation coefficients which were then compared.

Table 6. Correlations, Factors, and School Outcome Indicators.

Factor	ACT composite	End-of-course pass rate	Graduation rate	Poverty index
Combined (N = 183)				
Home–school relations	.46*	.49*	.44*	−.41*
School safety	.51*	.59*	.38*	−.47*
Physical environment	.45*	.51*	.37*	−.39*
Learning environment	.31*	.40*	.30*	−.26*
Teacher (N = 183)				
Home–school relations	.71*	.73*	.61*	−.69*
School safety	.42*	.57*	.35*	−.42*
Physical environment	.36*	.42*	.26*	−.31*
Learning environment	.54*	.68*	.37	−.53*
Student (N = 192)				
Home–school relations	.24*	.20*	−.04	−.18*
School safety	.50*	.50*	.06	−.46*
Physical environment	.42*	.43*	.07	−.38*
Learning environment	.20*	.20*	−.08	−.14*

*$p < .05$.

Looking at the combined data set, ACT scores, EOC pass rate, and poverty index scores exhibited moderate correlation values with factors. The strongest relations were observed between outcomes and School Safety (see Table 6), with the strongest correlation between EOC pass rate and the School Safety factor (.59).

The correlations between factors and school outcome indicators using teacher (common and unique) item factor scores are also presented in Table 6. All factor–outcome correlations were significant, with the exception of that between Graduation Rate and Learning Environment. Again, the highest significant correlations fell under the EOC pass rate outcome, ranging from .42 to .73. Indicators for the ACT composite and Poverty Index performed comparably across factors. Graduation Rate had the lowest significant correlation at .26 with Physical Environment.

For common and unique student survey items, relationships were much weaker than those for teachers (see Table 6). The strongest factor–outcome relationships were between School Safety and the outcomes for ACT composite (.50) and EOC pass rate (.50). Furthermore, Graduation Rate outcome was not significantly related to any student survey factors.

Students also exhibited stronger values than teachers on 5 of the 16 statistically significant correlations, especially between the poverty index and factor scores on both Learning Environment and Home–School Relations (.39 and .51 difference, respectively). Minimal differences ranging from .01 to .08 were noted for the three remaining factor–outcome correlation values. Factor–outcome correlation values for teachers, however, were very high for both Home–School Relations and Learning Environment factors and most school outcome indicators. Most factor–outcome correlation values were also higher under the School Safety and Physical Environment factors for teachers compared with students.

Together, these results show that the relationships between factor scores and school outcomes are much stronger for teachers compared with students and the combined factor scores. This suggests that the combination of factor scores may paint an inaccurate picture of school climate for each group, where teacher relationships are underestimated and student relationships are overestimated. Thus, robust, combined factor scores cannot be recommended in this case.

Discussion

The IDA framework expands the breadth and depth of a chosen construct via increased sample heterogeneity and size, improved capacity to integrate and replicate findings, and a better understanding of how constructs evolve over time (Bainter & Curran, 2015). It is this expansion that makes IDA a unique and feasible framework for use in an applied setting. The purpose of this study was to examine pooled school climate data from teacher and student school climate surveys using the IDA framework.

Our goal with this article was to provide an instructional approach for researchers interested in conducting IDA. Because our data were categorized by a single covariate, a large number of items, a high level of missingness, and more factors than what were presented in the existing literature, some changes were made to the IDA and MNLFA procedures set by Curran et al. (2014) and Bauer (2017). Curran et al. (2014) were able to bypass some of the aforementioned difficulties by using a unidimensional model that contained a small number of pooled items present in both studies, though, their study did contain multiple covariates. Bauer (2017) presented an example of a multidimensional model in his application of MNLFA procedures, but all items were dichotomized so that models could be fit using logistic specification and allow for moderated *nonlinear* factor analysis.

Differences between the current study and previous applications of IDA prompted our work to be largely exploratory in nature. The current study uses ordinal data given the use of Likert-type scale responses from teacher and student surveys. Furthermore, we encountered a large amount of missingness in our data which precluded the use of a categorical estimator (WLSMV). A robust maximum likelihood estimator (MLR) would have been a viable option if our data did not possess both nonnormality and missingness. Therefore, the data were treated as continuous, prompting a moderated *linear* factor analysis.

These modifications make our model more complex in some ways and less complex in others. While these decisions may not fully align with the methods adopted by Curran and Hussong (2009), Curran et al. (2014), and Bauer (2017), we believe them to be a novel attempt to integrate this method into practice. Future examination of IDA and MNLFA performance under more complex models could prove beneficial to researchers.

While the current study applied the IDA framework to school climate, the practical applications of IDA extend far beyond a single construct and may provide a more holistic approach to survey research. One purpose of applying IDA to school climate surveys was to create commensurate factor scores for student and teacher perceptions of climate. Unique group factor scores have provided the state of South Carolina with valuable insights on school climate and progress on accountability measures. However, in an attempt to combine teacher and student factor scores to predict school-level outcomes, we found that relationships between factor scores and outcome indicators were notably stronger for teachers than for students or pooled data.

More specifically, the combination of lower, negative factor–outcome correlations for students with the stronger relationships for teachers resulted in more moderate relationships between factor scores and outcome indicators across shared teacher–student survey items. This suggests that relationships between factor scores and outcomes are much stronger for teachers compared with students and that combining student and teacher factor scores may suppress the relationship between teacher perceptions of school climate and school outcomes.

These findings support the conclusion that combined factor scores may not be warranted in this case of school climate. While a robust combined score is feasible, information that helps us understand each group's perception of school climate is lost in the process. Combined scores should be looked at in addition to, not in place of, unique scores. Future research should investigate combined school factor scores in relation to a cluster analysis.

In addition to the issues involving the relationships/correlations, creating commensurate factor scores may prove problematic as a result of DIF. Statistically significant DIF was present across factor means and variances, as well as item intercepts and loadings. When creating factor scores, item invariance across groups is required, as the operationalization of theoretical constructs should be consistent across student and teacher surveys (Bauer & Hussong, 2009). The presence of DIF indicates that the constructs manifest differently across teachers and students and suggests that factor scores should be reported separately for each group. Thus, our ''DIF-detection strategy may have been somewhat overly inclusive'' (Bauer, 2017, p. 517), resulting in increased measure variance and the noncommensurate nature of student and teacher factor scores. This result may have been due to our large sample size (n = 39,737).

Due to the aforementioned findings, alternative strategies may need to be considered to improve the accuracy of DIF detection (e.g., logistic regression methods; Acar, 2011). If one is uncertain as to the similarity of the selected samples/groups, equivalence testing can be used to verify whether model parameters should be considered equal or invariant (see, Marcoulides & Yuan, 2017; Yuan, Chan, Marcoulides, & Bentler, 2016). Equivalence testing allows researchers to control the size of misspecification and to obtain corresponding, adjusted cutoff values of fit indices such as RMSEA and CFI (Yuan & Chan, 2016). We recognize that additional steps may be undertaken if researchers are not comfortable with the equality of their samples, and equivalence testing can be a viable option in such circumstances where one wishes to confirm the results of more conventional criteria typically used to assess invariance.

The time-intensive nature of IDA procedures was also a limitation. The final MNLFA model took approximately 35 hours to converge using *Mplus* software. Researchers who lack the time or operational capacity to handle the computational rigor of MNLFA procedures should proceed with prudence.

While IDA and MNLFA procedures provide a robust investigation method to show the structure of school climate surveys across students and teachers, it is our opinion that this method should complement and not replace individual-group reporting of school climate data. Teachers and students both make unique contributions to our understanding of school climate due to their different experiences and perceptions. For example, items about teacher working conditions were eliminated from the current analysis as they were not applicable to students. While this exclusion helped us compare students and teachers on similar factors, it also resulted in the loss of important information that can heavily impact teachers' perceptions of school climate. Participant-relevant instruments are more likely to capture these differences.

The dearth of applied research involving IDA leaves ample room for further exploration and analysis of its procedures and associated limitations. This is especially true as it relates to the procedure's effectiveness with complex models. Applications of IDA to fields outside of the public health field and using categorical data would give researchers and practitioners a broader view of how the procedure can be effectively applied across disciplines. Increased applications may prompt methodological advancements and more widespread and proficient use.

Within the field of survey research, IDA may enable researchers to more easily and affordably advance their research interests and developments in their respective fields. Researchers could cast a wider net and compare findings from studies that may not be feasible to complete due to limitations like time or funding. Given its flexibility with analyzing data from multiple sources, IDA should still be considered as an alternative data analysis approach with potential benefits for survey researchers.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Kathleen V. McGrath https://orcid.org/0000-0001-8670-2457

References

Acar, T. (2011). Sample size in differential item functioning: An application of Hierarchical linear modeling. *Educational Sciences: Theory and Practice*, *11*, 284-288. Retrieved from https://files.eric.ed.gov/fulltext/EJ919902.pdf

Asparouhov, T., & Muthén, B. (2010, August 14). *Weighted least squares estimation with missing data* (Technical report). Retrieved from http://www.statmodel.com/download/GstrucMissingRevision.pdf

Bainter, S. A., & Curran, P. J. (2015). Advantages of integrative data analysis for developmental research. *Journal of Cognition and Development*, *16*, 1-10. doi: 10.1080/15248372.2013.871721

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*, *22*, 507-526.

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*, *14*, 101-125. doi:10.1037/a0015583

Carlson, M., & Miller, N. (1987). Explanation of the relation between negative mood and helping. *Psychological Bulletin*, *102*, 91-108.

Cialdini, R. B., & Fultz, J. (1990). Interpreting the negative mood-helping literature via ''mega''-analysis: A contrary view. *Psychological Bulletin*, *107*, 210-214.

Crane, P. K., van Belle, G., & Larson, E. B. (2004) Test bias in a cognitive test: Differential item functioning in the CASI. *Statistics in Medicine*, *23*, 241-256.

Curran, P. J., Cole, V., Giordano, M., Georgeson, A. R., Hussong, A. M., & Bauer, D. J. (2018). Advancing the study of adolescent substance use through the use of integrative data analysis. *Evaluation and the Health Professions*, *41*, 216-245.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*, 81-100.

Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., & Zucker, R. A. (2008). Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. *Developmental Psychology*, *44*, 365-380.

Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A. B., Chassin, L., & . . . Zucker, R. A. (2014). A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. *Multivariate Behavioral Research*, *49*, 214-231. doi:10.1080/00273171.2014.889594

DiStefano, C., Mindrila, D., & Monrad, D. M. (2013). Investigating factorial invariance of teacher climate factors across school organizational levels. In M. S. Khine (Ed.), *Application of structural equation modeling in educational research and practice* (pp. 257-275). Rotterdam, Netherlands: Sense.

DiStefano, C., Monrad, D. M., May, R. J., McGuiness, P., & Dickenson, T. (2007, April). *Using school climate surveys to categorize schools and examine relationships with school achievement*. *Paper presented at the meeting of the American Educational Research Association*, Chicago, IL.

Ene, M., Leighton, E. A., Guo, Z., McGrath, K. V., DiStefano, C, & Monrad, D. M. (2016, April). *Relationships between school climate and indicators of school effectiveness: A multilevel investigation*. *Paper presented at the meeting of the American Educational Research Association*, Washington, DC.

Finney, S. J., & DiStefano, C. (2013). Non-normal and categorical data in structural equation models. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 439-492). Charlotte, NC: Information Age.

Gareau, S., May, R. J., Mindrila, D., Ishikawa, T., DiStefano, C., Monrad, D. M., & Price, K. (2010, April). *The relationship between school climate and school performance*. *Paper presented at the meeting of the American Educational Research Association*, Denver, CO, USA.

Glass., G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3-8.

Greenberg, E. (2004). *Climates for learning*. *Paper presented at the annual meeting of the American Educational Research Association*, San Diego, CA.

Hofer, S. M., & Piccinin, A. M. (2009). Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychological Methods*, *14*, 150-164.

Hox, J., & Bechger, T. (1995). Comparing and combining different approaches to the multitrait-multimethod matrix. In J. J. Hox, G. J. Mellenbergh, & P. G. Swanborn (Eds.) (Eds.), *Facet theory: Analysis and design* (pp. 135-144). Zeist, Netherlands: SETOS.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Hussong, A. M., Flora, D. B., Curran, P. J., Chassin, L. A., & Zucker, R. A. (2008). Defining risk heterogeneity for internalizing symptoms among children of alcoholic parents. *Development and Psychopathology*, *20*, 165-193.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Lee, V. E., & Burkham, D. T. (1996, February). Gender differences in middle-grade science achievement: Subject, domain, ability level, and course emphasis. *Science Edition*, *80*, 613-650.

Ma, X., & MacMillan, R. B. (1999). Influences of workplace conditions on teachers' job satisfaction. *Journal of Educational Research*, *93*, 39-47.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130-149.

Marcoulides, K. M., & Grimm, K. J. (2017). Data integration approaches to longitudinal growth modeling. *Educational and Psychological Measurement*, *77*, 971-989.

Marcoulides, K. M., & Yuan, K.-H. (2017). New ways to evaluate goodness of fit: A note on using equivalence testing to assess structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*, 148-153.

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal wave data with multiple samples and changing scales of measurement. *Psychological Methods*, *14*, 126-149.

McArdle, J. J., & Horn, J. L. (2002, September–October). *The benefits and limitations of mega-analysis with illustrations for the WAIS*. *Paper presented at the International Meeting of CODATA*, Montreal, Quebec, Canada.

Monrad, D. M., May, R. J., DiStefano, C., Smith, J., Gay, J., Mindrila, D., & . . . Rawls, A. (2008, April). *Parent, student, and teacher perceptions of school climate: Investigations across organizational levels*. *Paper presented at the meeting of the American Educational Research Association*, New York, NY.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Mutheä n & Mutheän.

Muthén, L. K., & Muthén, B. O. (2017). Mplus [Computer software]. Retrieved from https://www.statmodel.com/index.shtml

Muthén, L. K., & Muthén, B. O. (2019). *Mplus FAQs: Factor scores*. Retrieved from https://www.statmodel.com/faq.shtml

Penuel, W., Meyer, E., & Valladares, M. R. (2016, November 15). *Making the most of the Every Student Succeeds Act (ESSA): Helping states focus on school equity, quality and climate* (Policy Brief). Retrieved from http://nepc.colorado.edu/publication/ESSA

Rässler, S. (2003). A non-iterative Bayesian approach to statistical matching. *Statistica Neerlandica*, *57*, 58-74.

Stewart, E. B. (2007, June). School structural characteristics, student effort, peer associations, and parental involvement: The influence of school- and individual-level factors on academic achievement. *Education and Urban Society*, *40*, 179-204.

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

Wang, M. T., & Degol, J. L. (2016). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, *28*, 315-352. doi: 10.1007/s10648-015-9319-1

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 56-75). Newbury Park, CA: Sage.

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*, 10-18. doi:10.111/j.1750-8606.2009.00110.x

Yuan, K.-H., & Chan, W. (2016). Measurement invariance via multigroup SEM: Issues and solutions with chi-square difference tests. *Psychological Methods*, *21*, 405-426.

Yuan, K.-H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2016). Assessing structural equation models by equivalence testing with adjusted fit indices. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 319-330.