Chapter 3

Methodology

The methodology for the completion of this research is included in this chapter. This includes a review of participants, procedures, and analysis completed. A detailed depiction of participating students, teachers and schools is presented. In addition, the specific analytic procedures utilized to assess the different aspects of validity are discussed. The theoretical dimensional models and the final dimensional model are presented as well as a discussion of model identification, estimation and fit indices. Subsequently the remaining aspects of validity are described as they pertain to the evaluation of the SLEI measures. This includes a discussion of the applicable measurement model, rating scale analysis, item technical quality, precision and replicability and group differences.

Participants

Participants were identified through convenience sampling through university-school partnerships with the Fralin Biotechnology Center at Virginia Tech. Teachers were selected from each program based on several characteristics. These characteristics included location of school, type of academic program, and length of partnership with the University (Table 5). A total of 355 valid responses were received.

Table 5

SLEI Participants

Participant Group	Number of Teachers	Number of Students	Location
Partnership for Research in Plants	4	127	Arizona, Missouri and Virginia
Biotech-in-the –Box [DNA kit]	7	228	Virginia
Total	11	355	

Two programs conducted through the university's outreach effort were utilized to identify participating teachers. The Partnership for Research in Plants (PREP) is a laboratory based learning program. One hundred twenty seven students participated from this program. These students are from high schools in Virginia, Arizona and Missouri. High school students in Virginia utilizing biotechnology research projects from the Biotech-in-a-box kits also completed the instrument. This included 228 students. The course enrollment from both programs consisted of students in biology, agricultural science, and biotechnology in both regular and advanced classes in public and private schools. Teachers administered the survey to all of their students that participated in the program affiliated with the university, once parental consent and student assent had been obtained.

School size ranged from 300 to 4,000 students. Schools were located in a wide array of locations according to the National Center for Educational Statistics (NCES, 2007). These locations included rural, urban fringe (small, medium and large city), and small and large towns. The proportions of students receiving free or reduced price lunches at each school ranged from

3% to 40%. The ethnic minority population of the schools ranged from 10% to 40%. A 3% migrant population at one school was also included. Table 6 summarizes the demographic characteristics of participants' schools.

Table 6

Participating School Characteristics

Characteristic	Range
School Size	300 to 4,000 students
Location	Rural
	Urban fringe (small, medium and large city)
	Small and large town
	Military Base town
Free or reduced lunch	3% to 40%
Ethnic Minority	10% to 40%
Migrant	0 to 3%
Academic level	Advanced Placement, International Baccalaureate, Honors, Regular, Specialized Biotechnology

Procedures

Institutional review board approval for this research was obtained (a copy is located in the appendices). In addition, parental consent forms and student assent forms were approved by the Virginia Tech Institutional Review Board. Emails were sent to selected teachers from PREP and Biotech-in-a-Box. The emails provided brief information about the research efforts in educational research and requested the response of those teachers interested in participating. When responses were received, they were sent a package of materials for administration of the survey. This package included the following:

1. A brief memo of instruction to the teachers. This provided teachers with guidelines on consent form collection and survey administration. The memo also informed teachers about the return procedures and confidentiality of the survey data. Finally, the memo provided contact information for the researcher if the teacher had any questions.
2. An introductory letter to parents about the research.
3. Parental consent forms- one copy for their records and one copy to be returned.
4. Student assent forms.
5. The SLEI Survey - both forms attached together with the directions on the cover and the demographic information sheet appending the packet.

After consent and assent forms were collected from participating students, the teachers administered the surveys during their regular class sessions. The completion time for the survey was approximately 30 to 45 minutes. All materials were returned by mail to the researcher by the teachers for analyses after completion. In addition to the SLEI, the students were asked to complete a demographic information form (Appendix B). This form included questions related to gender, ethnicity, grade level, type of class, and length of class, school and teacher name. Confidentiality of students both at the school and individual level was maintained by the researcher.

Analyses

The following section details the methodology employed to examine the aspects of Messick's depiction of validity for the SLEI measures. Research questions 1 through 5 are addressed by each of these sections of analyses. Results from research questions 1 through 5a are discussed in Chapter 4. Results from research questions 5b and 5c are discussed in Chapter 5

Dimensionality

Research question 1(dimensionality), was examined by conducting dimensionality analyses of the SLEI responses through confirmatory factor analysis procedures in the LISREL program. There were three models that were selected a priori. This approach is seen as the strongest methodology for use in applications with confirmatory factor analysis (Bollen, 1989; Hoyle & Panter, 1995). Each of the theoretical models were supported by substantive research (Schumacker & Lomax, 2004).

The three models were selected because of existing research on classroom environments. Fraser identified the science lab as distinct from the broader learning environment (Fraser, McRobbie, & Giddings, 1993; B. J. Fraser, Giddings, & McRobbie, 1992; B. J. Fraser, Giddings, & McRobbie, 1995). His work was built on the earlier learning environment research of Moos (1987). The single dimensional model was examined as a baseline model to provide the broadest possible examination of the science laboratory learning environment construct. These models are further substantiated in Chapter 2.

In Fraser's model, there are five latent variables. These include student cohesiveness, open-endedness, integration, rule clarity and material environment. Each latent variable was designed to be measured by seven items on the "actual" form of the SLEI. Items 1, 6, 11, 21, 26 and 31 were designed to be indicators of student cohesiveness. Items 2, 7, 12, 17, 22, 27 and 32 were designed to measure open-endedness. Items 3, 8, 13, 18, 23, 28 and 33 were designed to measure the integration dimension. Items 4, 9, 14, 19, 24, 29 and 34 were designed to be indicators of rule clarity. The indicators for the material environment dimension included items 5, 10, 15, 20, 25, 30 and 35. The following table (Table 7) depicts the distribution of items by latent variable. In addition, reverse worded items are noted in the right most column.

Table 7

Item Mapping for Fraser Model

Actual	Student Cohesiveness	Open-endedness	Integration	Rule Clarity	Material Environment	Reverse worded items
1	x					
2		X				
3			x			x
4				x		
5					x	x
6	x					x
7		X				
8			x			x
9				x		x
10					x	
11	x					
12		X				
13			x			
14				x		
15					x	x
16	x					
17		X				
18			x			
19				x		
20					x	x
21	x					
22		X				
23			x			x
24				x		x
25					x	x
26	x					x
27		x				x
28			x			
29				x		
30					x	
31	x					
32		x				
33			x			x
34				x		
35					x	

In addition to the published SLEI model, two additional theoretical models were examined. These models include a unidimensional model and a three dimensional model. The unidimensional model depicts representation of the science laboratory as a single construct. This model defines the classroom environment as a single latent dimension with all items loading directly on it.

The three dimensional model is based on the research of Moos (1987) in psychosocial environment construction. His model includes three latent variables: relationship, personal and systems maintenance/ change. The relationship dimension is measured by items 1, 6, 11, 21, 26 and 31. Items 2, 3, 7, 8, 12, 13, 22, 23, 27, 28, 31 and 32 are indicators of the personal dimension. The indicators for the systems maintenance/ systems change dimension include items 4, 5, 9, 10, 14, 15, 19, 20, 24, 25, 29, 30, 34 and 35. The alignment between the Fraser and Moos dimensions is illustrated in table 8.

The model had 355 observations, with 35 different variables. In the final analysis, 22 variables were used after the reverse coded items were deleted. This resulted in 198 degrees of freedom, indicating an over identified model (Hoyle & Panter, 1995; Schumacker & Lomax, 2004). Model identification is completed to calculate the number of unique values in the sample as well as the number of free parameters in the models. The model should be over identified in order to be useful for analytical purposes (Bollen, 1989). An over identified model is one in which there are more elements in the variance-covariance matrix than parameters to be estimated. The degrees of freedom (d.f,= elements in variance-covariance matrix minus # of parameters to be estimated)are positive, not zero. Covariance matrices were used to estimate the different dimensional models. These were calculated using the JMP program ("JMP", 2006).

Multiple fit indices were used to evaluate the fit of the three models. Model–to-data fit for each of these models was examined based on three indices: Chi-square value, chi-squared divided by degrees of freedom (DF), and root mean square error of approximation (RMSEA). For the sake of interpretation, we chose to evaluate model-to-data fit prioritizing models with the following characteristics: (a) smaller chi-square values, (b) RMSEA values less than 0.05 (Schumacker & Lomax, 2004). Once models that exhibited suitable model-to-data fit were identified, the relative fit of these models was compared via the comparative fit index (CFI) and the normed fit index (NFI)—indices that compare the proposed model to a null model. Commonly accepted criterion values for these indices are those greater than 0.90 and optimally greater than 0.95. Model replicability was evaluated using the expected cross validation index (ECVI). The model with the smallest ECVI is considered the most optimal model from a replicability standpoint. Model parsimony was evaluated using the parsimonious fit index (PNFI) and consistent Akaike's information criterion (CAIC). These fit indices take the number of degrees of freedom in a model into consideration when examining parsimony. Ideal parsimonious conditions are present when there is a higher degree of fit with fewer degrees of freedom (James, Mulaik, & Brett, 1982). The model exhibiting higher values of PNFI and smaller values of CAIC are preferred. Furthermore, we examined the correlation matrix of the final dimensional model to verify that the measures for the various dimensions provided useful differentiation.

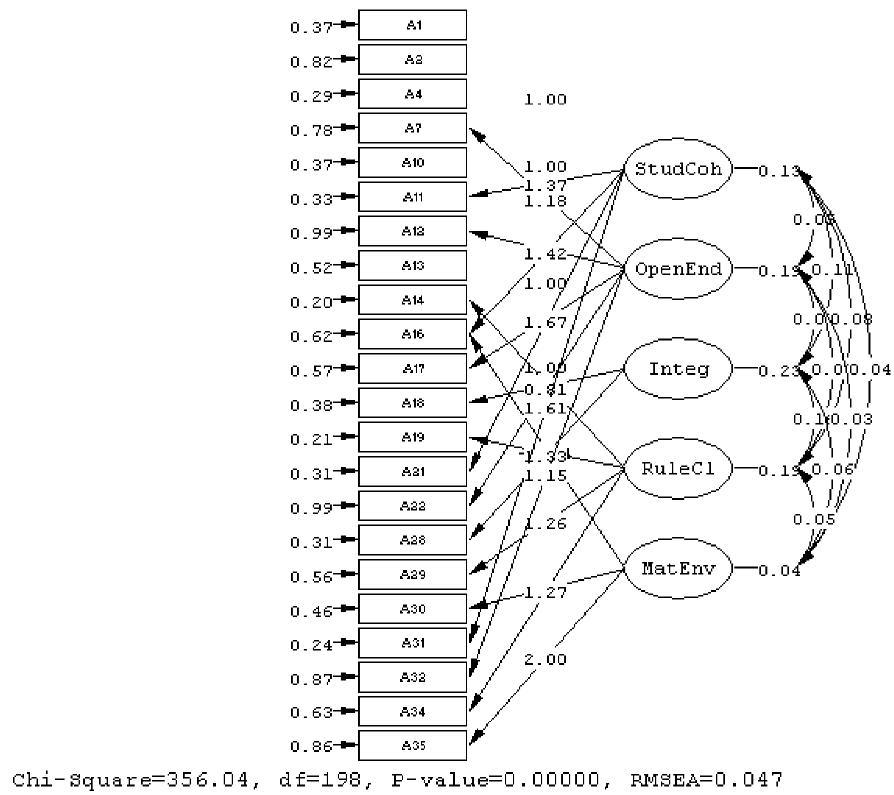

Figure 3 Final model illustration

Measurement Model

Research question 2 (Rating Scale Structure) addressed which measurement model and rating scale configuration best depicts the rating scale structure of these data. The data collected were assessed with the multi-dimensional random coefficients multinomial logit model (MRCMLM) (Adams, Wilson, & Wang, 1997) using the CONQUEST program(Wu, Adams, & Wilson, 1998). This model extends the Rasch one-dimensional model to include multiple traits present in responses to items. These multiple latent traits are defined by D and identify D-dimensions for the model based upon the analyses. The D-dimensions are arrayed in a vector of latent traits $\theta_n = [\theta_{n1}, \theta_{n2}, ... \theta_{nD}]$. A response in category k on item *i* of dimension d (D=1,2, …D) is scored b_{ikd}. The scores across D dimension for item *i* are represented by the vector $b_{ik} = [b_{ik1}, b_{ik2}, … b_{ikD}]$. Item validation (δ) is represented by a vector of *P* parameters $\delta_i = [\delta_{il}, \delta_{i2}, ... \delta_{ip}]$. Design vectors $a_{ik}=[i=1,…,I;\ k=1,…K_i]$ define linear combinations of these parameters that relate to empirical characteristics of the response categories of the item. Based upon the preceding definitions, the probability of a response in category *k* for item *i* is represented as

$$\pi_{nik} = \frac{\exp(b_{ik}\theta + a_{ik}\delta)}{\sum_{k=1}^{K_i}\exp(b_{ik}\theta + a_{ik}\delta)}$$ (Adams, Wilson, & Wang, 1997).

Rating Scale Analysis

Research question 2 addressed rating scale structure utilized by the students to determine if it was used as designed by the authors. In addition, question 2 addressed how well the rating scale met the guidelines as outlined by Linacre (2002). This question concerns the substantive aspect of validity. Linacre defines eight criteria for examining the optimization of rating scale category effectiveness. There should be at least ten observations per category to ensure a proportional distribution. The observation distribution should be fairly normal and unimodal. The rating structure should progress in a positive fashion where the ICC curves increase monotonically and no category should be under or overlapped by another category; average measures should advance monotonically. The mean square statistics should be examined to ensure that the observed and expected response rates behave similarly, and that there are no extreme values (>2.0). The minimum and maximum distances between rating scale reliability indices should be no lower than 1.4 and no higher than 5 (Linacre, 2004). Specifically, the analysis considered whether (a) rating categories were used with sufficient frequency (i.e., at least 10 observations per category), (b) the distribution of observation within each category was unimodal, (c) the rating scale thresholds increased in value with the rating categories, (d) the average measures associated with each rating category increased monotonically, (e) the mean square statistics associated with each category were reasonable (i.e., less than 2.0), and (f) the minimum and maximum distances between rating scale thresholds was greater than 1.4 and less than 5, respectively.

Item Technical Quality

Research question 3 (Item Quality) addressed the technical quality of the items. This concerns the content aspect of validity. Two forms of analyses were used to examine item technical quality. The point measure correlation measures the degree to which the score on one item is consistent with the scores on the remaining items. For purposes of this research, the minimum point measure correlation should be at least .30 (Wolfe & Smith, 2007b). If items do not meet the minimum threshold correlation, they will be highlighted in the presentation of results. The intent of this research was not to create a new form of the instrument, but to identify any potential validity issues and concerns. The standardized weighted mean square fit indices were considered, and items were flagged if the value of this index was greater than 2.00. Flagged items were examined for potential problems.

Precision and Replicability

Research question 4 (Reliability) investigated the generalizability aspect of validity. Generalizability was important to examine in order to understand whether the score interpretation was limited to the sample of assessed perceptions or was more broadly generalizable to the construct of science laboratory environments. Person separation reliability coefficients and precision indices were calculated for the five subscales of the SLEI to evaluate to reliability of the rank ordering of persons and the stability of the person measure estimates (Smith, 2001). These indices examined the extent that the observed responses for an individual were in concordance with the theoretical or predicted model. The reliability of separation index illustrated the degree to which each scale produced internally consistent measures. (Smith, 2001; Wolfe & Smith, 2007b).

Group Differences

Research question 5 (Group Differences) assessed any statistically significant differences in predicted response rates based on age, gender, ethnicity, or type of academic program. This external aspect of validity was reviewed through the examination of the means and standard deviations for individuals and groups for the five subscales of the SLEI to determine if there were any statistically significant differences. Expected response rates across these subgroups were reviewed to determine if there were any meaningful differences.