Chapter 4, Validation Article:

Assessing Student Perceptions of High School Science Classroom Environments: A Validation Study

Christine D. Luketic

Edward W. Wolfe

Kusum Singh

Erin Dolan

Virginia Tech

Blacksburg, Virginia 24061

Abstract

Measures from the Science Laboratory Environment Inventory (SLEI), an existing classroom environment measurement instrument, were examined in this validation study for the purpose of providing up-to-date norms and validation evidence for a U.S. secondary school population. Our structural equation modeling analyses revealed that a multi-dimensional model encompassing five distinct factors and excluding negatively-worded items best characterized the SLEI measures. Multidimensional measures created by scaling the data to the Multidimensional Random Coefficients Multinomial Logit (Rasch) Model exhibited suitable rating scale structure, item quality, and reliability of separation. In addition, comparison of gender and ethnicity groups revealed no differences in SLEI measures.

Science classroom environments have changed significantly in the past twenty years. These changes have been inspired, in a large part, by the National Science Education Standards, which provide recommendations for fundamental changes in pedagogical approaches to teaching science. These recommendations emphasize inquiry-based learning and prioritization of science as an area of study in our schools (NRC, 1996). A second impetus for change in the teaching of science has been the emphasis on educational accountability enacted in federal law. For example, No Child Left Behind ("No Child Left Behind", 2002) requires evaluation of achievement in public schools, which has led to increased attention on achievement of learning standards and, hence, the alignment of classroom learning and assessment with those standards.

Because of this transformation of the science curriculum and science pedagogy, it is important to ensure that we have valid measures for examining science learning and the effectiveness of science classroom instruction. In addition to achievement, researchers and science educators have emphasized that an important indicator of the effectiveness of the science classroom is the student perceptions of learning environments and experiences. Not only are these affective outcomes important in their own right, but several studies reveal that they are related to, and therefore may be integral to positive academic outcomes (Dorman & Adams, 2004; Fraser, 1998a; Newby & Fisher, 2000).

One instrument that measures student perceptions of the science learning environment, the *Science Laboratory Environment Inventory* (SLEI), was developed in the early 1990s for the purpose of assessing student perceptions of their science classroom laboratory experiences. However, it has been 15 years since the original validation evidence for the measures from the SLEI was collected. The changes in classroom practices, science curriculum, and the student population, along with the increased scrutiny resulting from accountability legislation, have rendered much of that evidence outdated. Hence, science education has changed sufficiently to warrant re-examination of validity evidence relating to SLEI measures.

The Science Laboratory Environment Inventory

The SLEI measures 5 aspects of the science laboratory learning environment: (a) student cohesiveness, (b) open-endedness, (c) integration, (d) rule clarity, and (e) material environment. There are two forms of the instrument. The “actual” form was designed to measure student perceptions of the experienced learning environment. Student perceptions of the favored learning environment are assessed through the “preferred” form (Fraser, 1998a, 1998b, 2002, 2007; Fraser, Giddings, & McRobbie, 1992; Fraser, McRobbie, & Giddings, 1993). Each form contains seven statements relating to each of the five aspects of the environment for a total of 35 items. The statements describe lab activities, rules, equipment and student to student and student to teacher interactions. Students respond by indicating how often these activities take place on a five-point scale ranging from “almost never” to “very often.” Thirteen of the items are reverse-scored.

The items of the instrument are based on a general learning theory specified by Moos that was later adapted to science learning contexts by Fraser, Giddings and McRobbie (1993). Moos developed a three-part psychosocial theory of learning environments that focuses on the relationships, personal development, and systems maintenance/change that arise in those environments (Dorman & Adams, 2004; Moos, 1987). The relationship dimension of this model refers to the nature and intensity of personal relationships. The personal development dimension refers to the path through which knowledge development progresses. The system maintenance and change dimension refers to the orderliness, clarity, control, and responsiveness to change in the environment.

Fraser built on Moos' model, adapting it to describe classroom learning environments (initially) and (later) science learning environments. Fraser's work with McRobbie and Giddings identified five aspects of this environment—the aspects around which the SLEI items were developed. Student cohesiveness describes how well students know each other, work well together, and support one another. Open-endedness refers to students' opportunities to design their own research and pursue individual interests to enhance their personal construction of scientific knowledge. The integration dimension characterizes how lab activities are connected to theoretical material taught in the lecture portion of the science classroom. Rule clarity is defined by how clearly structure and expectations are communicated and implemented in the lab. Material environment describes students' perceptions of the adequacy of their lab materials and equipment.

The authors of the SLEI conducted investigations of the psychometric properties of the instrument when it was initially developed (Fraser, McRobbie, & Giddings, 1993). Coefficient alpha for the subscales ranged from .75 to .83, although the values varied considerably between international populations (e.g., reliabilities ranged from a low of .49 in Nigeria on the open-endedness scale to a high of .89 in Israel on the integration scale). Subscale correlations ranged from a low of .07 (between open-endedness and the other scales) to a high of .37 (between integration and the other scales). In addition, evidence of the external aspect of validity on these measures was collected through analysis of the responses within and between classrooms. Specifically, analyses revealed that students within each class held similar perceptions of the learning environment while between-classroom perceptions were differentiated.

Prior research concerning the substantive aspect of validity for the SLEI measures has suggested that the dimensional structure of the instrument is 5 dimensional including 13 negatively worded items (Fraser, McRobbie, & Giddings, 1993). Negatively worded items have been often used and discussed in instrument design as a way to modify respondents' behaviors and detect non-attending behaviors. Adoption of negatively worded items is claimed to reduce instances of response set (Barnette, 2000; Bergstrom, 1998; Marsh, 1996; Schriesheim & Hill, 1981; Yamaguchi, 1997). However, responses to negatively worded items do not necessarily represent polar opposite responses from positive items. Research has shown that negatively worded items tend to cluster as if they constituted a unique dimension unto themselves (Schriesheim, Eisenbach, & Hill, 1991), which may adversely affect internal consistency of the measures (Chamberlain & Cummings, 1984; Marsh, 1996; Schriesheim, Eisenbach, & Hill, 1991). As a result, several researchers have recommended that the practice of using of negative items to control response sets should be eliminated (Chamberlain & Cummings, 1984; Schriesheim, Eisenbach, & Hill, 1991; Schriesheim & Hill, 1981; Yamaguchi, 1997).

Additional validation evidence that could be developed for the SLEI measures, evidence that has not been provided to date, could come from comparing the perceptions of different ethnicities or genders. Concerning ethnicity differences in learning environment perceptions, existing research has not utilized the SLEI. In a study that employed two related measures of perception of learning environments among students in the Pacific islands, the Questionnaire on Teacher Interaction and the College and University Classroom Environment Inventory, researchers found minimal evidence of differences in perception of teacher student interactions based on ethnicity for this population (Coll, Taylor, & Fisher, 2002). On the other hand, prior research has suggested that gender differences may influence student perceptions of the learning environment. In one study, researchers compared Australian male and female responses to the Learning Process Questionnaire and determined that females were more positive about their learning environments, although the observed effect sizes were small (Dart et al., 1999). In the Coll, Taylor, and Fisher (2002) study, Pacific island university females also reported more positive perceptions of the learning environment than did males. Research conducted with students from Singapore using the SLEI revealed statistically significant differences for gender in the dimensions of student cohesiveness, rule clarity, and material environment. The researchers found that gifted males and females perceive open-endedness similarly but that non-gifted females provided more positive depictions of classroom open-endedness than did males. Non-gifted males students were, on the other hand, more positive about the material environment than were the non-gifted female students (Lang, Wong, & Fraser, 2005).

Unfortunately, the evidence concerning the reliability and validity of the SLEI measures is somewhat dated and represents a range of performance scores from multiple countries without presenting a cohesive body of evidence for any one of those countries. Specifically the existing evidence makes it difficult to interpret measures from the instrument for research based on U.S. secondary students in light of some inconsistent results that have been obtained concerning basic psychometric properties of the measures, such as reliability. In addition, the evidence to date is scant when held to current standards for validity and reliability (Wolfe & Smith, 2007a, 2007b). For example, research concerning the instrument measures has not focused on at least two critically important validity aspects. First, no attention has been directed toward the quality of the items, Messick's (1995) content aspect of validity. Second, no attention has been directed toward the quality of the rating scale that is used to elicit responses from students, Messick's structural aspect of validity. In addition, prior studies of the SLEI have only addressed the external aspect of validity tangentially. Specifically, that work has only verified that different classrooms have different perceived learning environment characteristics. The purpose of this article is to provide up-to-date norms for the measures of the SLEI for populations of U.S. secondary students and to present updated and more complete evidence of the validity of SLEI measures for depicting U.S. student perceptions of the science learning environment.

The following research questions were considered in this study:

- Research Question 1 (Dimensionality): What dimensional structure best depicts student attitudes toward their science laboratory environment?
- Research Question 2 (Rating Scale Structure): Which measurement model and rating scale configuration best depicts the rating scale structure of these data?
- Research Question 3 (Item Quality): What is the technical quality of the items?
- Research Question 4 (Reliability): What is the reliability of the subscale measures?
- Research Question 5 (Group Differences): Are there gender and ethnicity group differences across the measures?

Method

Participants

Classrooms participating in a university-school partnership focusing on authentic science inquiry were selected purposively in order to maximize sample diversity. Participating schools were chosen to ensure that there were responses from different grade levels, high achieving and regular classrooms, as well as different community settings. Specifically, schools were chosen from three states (Virginia, Arizona, and Missouri) to represent private and public schools as well as the duration of the partnership with the university (1 to 4 years). The SLEI was distributed to 900 students in 17 schools. A total of 355 students from 11 schools provided valid responses to the questionnaire (a 39% response rate). Several teachers who initially agreed to participate were unable to do so due to scheduling problems, central office administrative issues, and delays in obtaining parental consent. In addition, 48 responses to the instrument could not be utilized due to a lack of student/parent consent.

In the obtained sample, student to teacher ratios ranged from 12.4 to 20.5 students per teacher. School size ranged from 300 to 4,000 students, and school settings included rural, urban fringe (small, medium and large city), and small and large towns. The proportions of students receiving free or reduced lunch at each school ranged from 3% to 40%, and the racial minority population of the schools ranged from 10% to 40% (NCES, 2007). The sample of 355 students was predominantly female (65%) and European American (68%). Thirty-five percent (n= 123) of the students were males and 65% (n=226) were females.

Administration Procedures

Students shared their perceptions of the science classroom laboratory environment through completion of the SLEI from late December 2006 through April 2007. Teachers administered the SLEI questionnaires during regular class sessions. The completion time for the instrument was approximately 30 minutes. In addition to the SLEI, students completed a one-page demographics questionnaire, including grade in school, academic program, gender, and ethnicity.

Analyses

Dimensionality

Research Question 1 (Dimensionality) was addressed via confirmatory factor analyses conducted within a structural equation modeling framework. 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.

Three models were considered in our analysis of dimensionality: (a) a unidimensional model, (b) a three dimensional model based on Moo's taxonomy of learning environments, and (c) a five dimensional model based on Fraser's structure of science laboratory learning environments. Initially, model-to-data fit was not adequate for any of these models, so a fourth, non-substantive, model was fit to the data, differentiating negatively and positively worded items. Based on this analysis, negatively worded items were removed from the data analysis, and the fit of the original three models was reassessed, and the remaining analyses were conducted on the best fitting model identified in those analyses.

Once an appropriate model was selected based on the dimensionality analyses, data were scaled to the selected model via the multi-dimensional random coefficients multinomial logit model (MRCMLM) (Adams, Wilson, & Wang, 1997). The MRCMLM extends the family of Rasch models to measurement contexts in which multiple traits contribute to the item responses. The *D* latent traits are arrayed in a vector, $\theta_n = \left[\theta_{n1}, \theta_{n2}, ..\theta_{nD}\right]$. A response in category *k* on item *i* of dimension *d* is scored b_{ikd}. The scores across *D* dimension for item *i* are represented by the vector $b_{ik} = [b_{ik1}, b_{ik2}, \ldots b_{ikD}]$. Item endorseabilities (δ) are represented by a vector of *P* parameters, $\delta_i = \left[\delta_{il}, \delta_{i2}, ..\delta_{ip}\right]$. Design vectors, $a_{ik} = [i=1,\ldots,I;\ k=1,\ldots K_i]$, define linear combinations of item 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 $\pi_{nik} = \dfrac{\exp(b_{ik}\theta + a_{ik}\delta)}{\sum_{k=1}^{K_i}\exp(b_{ik}\theta + a_{ik}\delta)}$ (Adams, Wilson, & Wang, 1997).

Rating Scale Structure

Two analyses addressed Research Question 2 (Rating Scale Structure) First, parameters for both a rating scale and a partial credit version of the MRCMLM were estimated, and Bayesian Information Criterion (BIC) transformations of the deviance (G^2) statistics for these two models were compared to determine whether the rating scale versus the partial credit model best explained the observed rating scale use. Because the chi-squared statistic is sensitive when sample size is large, we also examined the threshold estimates from the partial credit scaling across items. Second, the adequacy of the ratings assigned by the students was evaluated based on a subset of guidelines recommended by Linacre (Linacre, 2004). Specifically, we 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 Quality

To address Research Question 3 (Item Quality), we considered two indicators of the technical quality of the items. First, the item score-theta correlation (AKA, point-measure correlation) was examined, and items were flagged if the value of this index was less than .30. Second, 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.

Reliability

Analyses relating to Research Question 4 (Reliability) focused on the reliability of separation index of theta for each dimension modeled in the MRCMLM. This index depicts the estimated proportion of observed variance that is true variance, and is interpreted in a manner comparable to coefficient alpha.

Group Differences

We addressed Research Question 5 (Group Differences) by determining whether gender or ethnicity groups exhibited different means on the five subscales of the SLEI. The ethnicity variable was divided into two groups: white/Caucasian and ethnic/racial minority, and we did not expect to find differences between groups. We did, however, expect to find differences based on gender based on previous studies.

Results

Dimensionality

The dimensional analyses are summarized in Table 1. First, note that the bottom row of this table summarizes the fit of the five-dimensional Fraser model with the 13 negative items included in the analysis. Comparison of fit indices (e.g., Chi-square, RMSEA) indicates a dramatic improvement in model-to-data fit when the negatively worded items are removed from the analysis. This was true for all three substantive models, so the top three data rows of Table 2 compare the confirmatory factor analysis results with the negatively worded items removed. Examination of the multiple fit indices across these three substantive models indicates that the Fraser five-dimensional model exhibited the best model-to-data fit (χ^2, χ^2/Df, and RMSEA), relative fit (CFI and NFI), replicability (ECVI), and model parsimony (PNFI). The lower diagonal of Table 3 indicates that the expected correlations between the latent variables measured by the SLEI are not redundant. Specifically, those correlations are estimated to range from .17 to .75, indicating these are distinct factors.

While additional evidence is typically reported in SEM-based confirmatory factory analysis concerning the adequacy of the measurement model, that evidence is comparable to, but less complete than, what is available from the multi-dimensional Rasch analyses. Hence, the remaining information reported here focuses on the results of the MRCMLM results.

Norms

Descriptive statistics for the scaled measures of each subscale of the SLEI are shown in Table 2. Specifically, the tables display the mean, standard deviation, and deciles for each subscale for the sample of 355 U.S. secondary students.

Rating Scale

Concerning Research Question 2 (Rating Scale Structure), the relative fit of the partial credit and rating scale versions of the MRCMLM configuration of Fraser's five-dimensional model, as measured by the BIC values, indicates a statistically significant improvement in fit of the partial credit model over the rating scale model, $\chi^2_{62} = 113.97$, $p < .0001$. However, comparison of the threshold estimates from the partial credit scaling across items indicated that application of the rating scale model to these data is reasonable. Table 4 summarizes the analysis of rating scale category used for the rating scale model. In general, Linacre's criteria were satisfied by the current five-point configuration of the rating scale. However, the thresholds for categories 2 ("Seldom") and 3 ("Sometimes") were fairly close together, which resulted in average student measures that did not increase on the Student Cohesiveness and Integration subscales.

Reliability & Item Quality

Table 5 summarizes the multidimensional student measure reliability coefficients and descriptive statistics relating to the item calibrations estimated for the MRCMLM, which address Research Questions 3 and 4 (Item Quality and Reliability). The reliability coefficients range from a low of .57 to a high of .77, suggesting that all of the scales except for Open-Endedness exhibit sufficient reliability for use in research settings. The point-measure correlations are reasonably high with only three items across the subscales exhibiting values less than .30. Scrutiny of these items indicated that their calibration values were relatively extreme, a condition that attenuates the value of the point-measure correlation. None of the items exhibited misfit.

Group Differences

Research Question 5 (Group Differences) was addressed by comparing the measures of ethnicity groups and gender. Table 6 presents the descriptive statistics for these comparisons. Overall, the differences between group means were very small for all subscales of the SLEI. None of these differences were statistically significant.

Discussion

Our results provide fairly positive validity evidence concerning the measures of the SLEI as indicators of student perceptions of the science laboratory environment. First, concerning Research Question 1 (Dimensionality), we found that a five dimensional model, removing negatively worded items, was the best model. Prior research has suggested that negatively worded items often do not function as intended, so this is not a surprising outcome (Marsh, 1996; Schriesheim & Hill, 1981; Yamaguchi, 1997). In addition, the moderate to low latent variable correlations suggest that the five-way distinction seems reasonable and that measures of these dimensions do not provide unnecessarily redundant information about student perceptions—an outcome that is consistent with the validation work completed on these measures in the early 1990's (Fraser, McRobbie, & Giddings, 1993). This fact is particularly important from the perspective of researchers or program evaluators who would like to depict science laboratory learning in a diagnostic manner because the relatively low correlations between the latent variables will allow for the development of profiles that indicate relative strengths and weaknesses of individual classrooms or programs.

Second, concerning Research Question 2 (Rating Scale Structure), our results suggest that the five-point rating scale structure used on the SLEI is not optimal and that, perhaps, a four-point scale may work better. There have been no prior analyses of the SLEI rating scale's function, so this constitutes a unique outcome. Future studies will be necessary to determine whether a four-point rating scale for the SLEI does, indeed, remedy the minor problems we identified concerning the proximity of the lower rating scale thresholds on all five of the SLEI subscales. Third, concerning Research Question 3 (Item Quality), our results suggest that the SLEI items work very well as indicators of student perceptions of the science laboratory environment. None of the items were flagged for serious concern, again, an outcome that is consistent with previous studies (Fraser, Giddings, & McRobbie, 1995; Fraser, McRobbie, & Giddings, 1993). However, those studies did not identify the negatively polarized items as being problematic, and our dimensionality analyses did indicate potential problems with those items.

Fourth, our results relating to Research Question 4 (Reliability) indicate that the multidimensional reliabilities are reasonably high and are consistent with Fraser's previous work (1992). However, we should note that our reliabilities were lower than those found in the earlier validation work even though we utilized a multidimensional scaling model. Perhaps this is the result of our sample being more homogeneous than samples in previous work, coupled with the fact that we eliminated 13 reverse worded items leaving several dimensions with low numbers of items.

Fifth, concerning Research Question 5 (Group Differences), we found no gender or ethnicity differences observed on any of the subscales. Research in Singapore utilizing the SLEI revealed gender differences in perception of the classroom environment (Lang, Wong, & Fraser, 2005), while there has been little research concerning the comparison of student perceptions of their classroom based on ethnicity for the measures of the SLEI. Although we expected to observe no ethnicity differences, the lack of gender differences on the SLEI was not consistent with our expectations and may be due to the socio-cultural differences between U.S. students and Singapore students.

Because this study focused on a group of students that was fairly small (N = 355) and because we have suggested minor modifications of the rating scale structure employed on the SLEI and removal of the negatively polarized items, future validation studies that evaluate a revised version of the SLEI should employ a larger sample of students. Although our sample contained a diverse representation of students across multiple ethnicities, many of the distinct groups did not have a sufficient number of responses for detailed ethnicity comparisons. An additional minor concern is that two dimensions are measured by only three or four items. The fact that we scaled our data with a multidimensional model allowed us to achieve higher levels of reliability than would be the case for true score or separate calibrations for the subscales. Hence, future researchers should either adopt a multidimensional scaling model or may want to consider developing additional items for these dimensions in order to increase the levels of reliability. Of course, the fact that our sample may be more homogeneous than a more general population means that reliability estimates for a more heterogeneous population will likely be higher than what was observed in this study.

Science teaching has changed significantly in the past 15 years. In addition to these different teaching approaches, students have also been learning differently. This validation study has provided additional and up-to-date evidence concerning the validity of SLEI measures. Coupled with the existing research revealing positive correlations between student perceptions of their learning environments and their academic achievement (Fraser, McRobbie, & Giddings, 1993), the outcomes reported here provide a foundation for future assessment of the relationship between classroom environment and student achievement. Researchers need clear evidence to substantiate these connections.

References

Adams, R., Wilson, M., & Wang, W. (1997). The Multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1-23.

Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: if you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement, 60*(3), 361-370.

Bergstrom, B. A. (1998). *Rating scale analysis: Gauging the impact of positively and negatively worded items.* Paper presented at the American Educational Research Association, San Diego.

Chamberlain, V. M., & Cummings, M. N. (1984). Development of an instructor/ course evaluation instrument. *College Student Journal, 18*, 246-250.

Coll, R. K., Taylor, N., & Fisher, D. L. (2002). An application of the questionnaire on teacher interaction and college and university classroom environment inventory in a multicultural tertiary context. *Research in Science and Technological Education, 20*(2), 165-183.

Dart, B., Burnett, P., Boulton-Lewis, G., Campbell, J., Smith, D., & McCrindle, A. (1999). Classroom learning environments and students' approaches to learning. *Learning Environments Research, 2*, 137-156.

Dorman, J., & Adams, J. (2004). Associations between students' perceptions of classroom environment and academic efficacy in Australian and British secondary schools. *Westminster Studies in Education, 27*(1), 69-85.

Fraser, B. J. (1998a). Classroom environment instruments: development, validity and applications. *Learning Environments Research, 1*, 7-33.

Fraser, B. J. (1998b). Science learning environments: Assessment, effects and determinants. In B. J. Fraser & K. G. Tobin (Eds.), *The international handbook of science education* (pp. 527-564). Dordrecht, The Netherlands: Kluwer.

Fraser, B. J. (2002). Learning environment research: Yesterday, today and tomorrow. In S. C. Goh & S. M. Khine (Eds.), *Studies in educational learning environments* (pp. 1-27). Singapore: World Scientific.

Fraser, B. J. (2007). Classroom Learning Environments. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 103-124). Mahwah, NJ: Lawrence Erlbaum Associates.

Fraser, B. J., Giddings, G. J., & McRobbie, C. J. (1992). *Science laboratory classroom environments: A cross-national study.* Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, Boston.

Fraser, B. J., Giddings, G. J., & McRobbie, C. J. (1995). Evolution and validation of a personal form of an instrument for assessing science laboratory classroom environments. *Journal of Research in Science Teaching, 32*, 399-422.

Fraser, B. J., McRobbie, C. J., & Giddings, G. J. (1993). Development and cross-national validation of a laboratory classroom environment instrument for senior high school science. *Science Education, 77*(1), 1-24.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills, CA: Sage.

Lang, Q. C., Wong, A. F. L., & Fraser, B. J. (2005). Student perceptions of chemistry laboratory learning environments, student teacher interactions, and attitudes in secondary school gifted education classes in Singapore. *Research in Science Education, 35*, 299-321.

Linacre, J. M. (2004). Rasch model estimation: Further Topics. In J. Everett V. Smith & R. M. Smith (Eds.), *Introduction to Rasch Measurement* (pp. 48-72). Maple Grove, MN: JAM Press.

Marsh, H. W. (1996). Positive and negative global self esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70*, 810-819.

Moos, R. H. (1987). *The social climate scales: A user's guide*. Palo Alto: Consulting Psychologists Press.

NCES. (2007). *Search for schools, colleges and libraries*. Retrieved. from www.nces.ed.gov/global locator/.

Newby, M., & Fisher, D. (2000). A model of the relationship between university computer laboratory environments and student outcomes. *Learning Environments Research, 3*(1), 51-66.

No Child Left Behind, Stat. 1425 (2002).

NRC (Ed.). (1996). *National science education standards*. Washington, DC: National Academy Press.

Schriesheim, C. A., Eisenbach, R. J., & Hill, K. D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement, 51*, 67-78.

Schriesheim, C. A., & Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement, 41*, 1101-1114.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah: Lawrence Erlbaum

Wolfe, E. W., & Smith, E. V., Jr. (2007a). Instrument development tools and activities for measure validation using Rasch models: Part I - instrument development tools. In E. V. J. Smith & R. Smith (Eds.), *Rasch Measurement: Advanced and Specialized Applications* (pp. 202-242). Maple Grove, MN: JAM Press.

Wolfe, E. W., & Smith, E. V., Jr. (2007b). Instrument development tools and activities for measure validation using Rasch models: Part II- validation activities. In E. V. J. Smith & R. Smith (Eds.), *Rasch Measurement: Advanced and Specialized Applications* (pp. 243-290). Maple Grove, MN: JAM Press.

Yamaguchi, J. (1997). Positive versus negative wording. *Rasch Measurement Transactions, 11*, 567.