10-8-2010

Analyzing Multilevel Data: An Empirical Comparison of Parameter Estimates of Hierarchical Linear Modeling and Ordinary Least Squares Regression

Louis M. Rocconi

To the University Council:

The Dissertation Committee for Louis M. Rocconi certifies that this is the final approved version of the following electronic dissertation: "Analyzing Multilevel Data: An Empirical Comparison of Parameter Estimates of Hierarchical Linear Modeling and Ordinary Least Squares Regression."

Corinna A. Ethington, Ph.D.
Major Professor

We have read this dissertation and recommend its acceptance:

John C. Smart, Ph.D.

Gary R. Pike, Ph.D.

Terry T. Ishitani, Ph.D.

Yonghong J. Xu, Ph.D.

Accepted for the Graduate Council:

Karen D. Weddle-West, Ph.D.
Vice Provost for Graduate Programs

ANALYZING MULTILEVEL DATA: AN EMPIRICAL COMPARISON OF PARAMETER ESTIMATES OF HIERARCHICAL LINEAR MODELING AND ORDINARY LEAST SQUARES REGRESSION

by

Louis M. Rocconi

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Educational Psychology and Research

The University of Memphis

December 2010

To my family

Dad, Mom, Dave, Kathleen, and Mary Anne

in appreciation of your love, support, and patience

ACKNOWLEDGEMENTS

This dissertation is the result of three years of doctoral study at the University of Memphis. I would like to express my sincere gratitude to all those who supported me throughout the process. First and foremost, I would like to thank my advisor, Dr. Corinna “Bunty” Ethington for her support, assistance, advice, and patience throughout my doctoral studies. Thank you, Bunty, for always encouraging me to do my best and for challenging me redo work that was not my best. I would also like to thank Dr. John Smart for encouraging me to enter the doctoral program and advising me throughout the program. John and Bunty have mentored me throughout my academic career and bestowed their expertise and passion for research upon me. The relationship we have developed throughout this journey has been one of mutual respect and friendship which I will always hold dear to my heart.

I would also like to thank my committee members, Dr. Yonghong Jade Xu, Dr. Terry Ishitani, and Dr. Gary Pike, who provided guidance and insight into my research. Thank you, Jade and Terry, for your years of advice and assistance throughout my doctoral program. Thanks also to Gary for assisting with obtaining the dataset for my dissertation and for the thoughtful and helpful comments throughout my dissertation process.

Thanks also to the helpful staff at The Indiana University Center for Postsecondary Research for assisting me with obtaining the data used in this project. NSSE data were used with permission from The Indiana University Center for Postsecondary Research.

Finally, I would like to thank Dr. Daniel Poje for his words of advice and encouragement throughout the process, for letting me complain to him when times were tough and for encouraging me to take time for myself. Dan has been a great support system and mentor to me in all my academic endeavors. Without the guidance from all of these people, I would likely have not have survived this dissertation project with my sanity.

ABSTRACT

Rocconi, Louis Marshall, Ph.D. The University of Memphis. December 2010. Analyzing Multilevel Data: An Empirical Comparison of Parameter Estimates of Hierarchical Linear Modeling and Ordinary Least Squares Regression. Major Professor: Corinna A. Ethington, Ph.D.

How college affects students is a central phenomenon of interest in higher education research. However, a major problem in assessing the influence of college on students is the methodological dilemmas due the multilevel nature of the majority of data used in such studies. Historically, higher education researchers have utilized the traditional linear model, ordinary least squares (OLS) regression, to aid in their investigation of the influence of college on students. This traditional approach ignores the multilevel nature of the data which can cause a multitude of conceptual and statistical problems. Therefore, a statistical technique, such as hierarchical linear modeling (HLM), that takes into account the multilevel nature of the organization of higher education is need. The purpose of this study is to determine whether conclusions regarding the influences on college seniors' critical thinking ability would differ depending upon the type of analysis, OLS regression or the more appropriate HLM analysis. In this study, the influences on seniors' critical thinking ability is examined three ways— (1) an OLS regression with the student as the unit of analysis, (2) an OLS regression with the institution as the unit of analysis, and (3) a three-level HLM with student attributes modeled at Level 1, characteristics of the major modeled at Level 2, and characteristics of the institution modeled at Level 3— in order to illustrate the differing conclusions one may come to depending upon the type of analysis chosen. Overall, evidence from this sample suggest that one would come to substantively different conclusions regarding the influences on students' perceived critical thinking ability depending upon the type of

analysis chosen, especially in regards to the effects of the institutional characteristics. Specially, the results from the institution-level OLS regression cannot be considered reliable. Findings from the institution-level OLS regression model differed substantially from the results of the other two analyses. The results from the student-level OLS regression analysis can only be partially trusted. The student-level OLS regression produced results comparable to the HLM estimates for the lower-level variables but substantively different results for the institutional characteristics. Thus, when institutional characteristics are of prime importance, one should perform an HLM analysis in order to be confident in the results obtained for the institutional effects.

TABLE OF CONTENTS

Chapter		Page
1	Introduction	1
	Brief History of HLM	7
	Robustness Issues with OLS Regression	9
	Robustness Issues with HLM	10
2	Methodology	14
	Sample	14
	The Model	15
	Variables	16
	Data Analysis	19
3	Results	27
	Student-level OLS Regression	27
	Institution-level OLS Regression	27
	Hierarchical Linear Modeling Analysis	30
4	Discussion	38
	OLS Regression with the Student as the Unit of Analysis	38
	A Comparison of the Two Regression Models	41
	Appropriateness of Hierarchical Linear Modeling	43
	HLM Estimates of Variance Components	46
	Comparison of HLM Fixed Effects to the two Regression Models	49
	Conclusion	55
References		62
Appendices		
A.	Items Comprising the Variables used in the Analyses and the Construction of the Scales	67
B.	List of Majors and Biglan Classification	71
C.	Correlations, Means, and Standard Deviations for Student-level OLS Regression	73
D.	Correlations, Means, and Standard Deviations for Institution-level OLS Regression	74

LIST OF TABLES

Table		Page
1	Student-level OLS Regression Results	28
2	Institution-level OLS Regression Results	29
3	HLM Estimation of Unconditional Model	31
4	HLM Estimation of Random Effects with Student-level Predictors (Level-1 Model)	33
5	HLM Estimation of Random Effects with Student-level and Major-level Predictors (Level-2 Model)	34
6	HLM Estimation of Random Effects with Predictors Modeled at all Three Levels (Level-3 Model)	35
7	HLM Estimation of Fixed Effects with Predictors Modeled at all Three Levels	37
8	Comparison of Results across Analyses	50

CHAPTER 1

INTRODUCTION

A central phenomenon of interest in higher education research is the impact of college on students. Scholars seek to understand how personal characteristics of students and aspects of their educational experiences influence students' academic learning and growth. This learning chiefly takes place in the organizational settings of institutions and features of these settings can have substantial influences on students' growth and development in college (Pascarella & Terenzini 1991, 2005; Toutkoushian & Smart, 2001). While it is widely recognized that institutional characteristics impact students' growth and development, a major problem in assessing that impact on students is the methodological dilemmas due to the multilevel character of the majority of data used in such studies.

Multilevel or hierarchical data are a common fixture in higher education. The classic example of multilevel data in higher education is students grouped or 'nested' within institutions. Multi-institutional datasets often contain variables that describe students as well as variables that describe institutions. For instance, data collected on college students may contain variables that describe students, such as interactions with faculty members and other students, experiences in coursework and extracurricular activities, as well as variables that describe institutions, such as sector, selectivity, and graduation rates. Additionally, even single-institution studies could have a hierarchical nature given the organization of postsecondary institutions. Students are nested in classes, majors, departments, and colleges or schools within an institution. Furthermore, it is common to find analyses with students nested within academic majors nested within institutions, where the individual, major, and institution are all the objects of interest and of observation. Despite the prevalence of hierarchical structures present in post-secondary educational research, past studies have often failed to address them adequately in the data analysis (Burstein, 1980; Ethington, 1997; Pascarella & Terenzini, 1991, 2005).

In his seminal critique of educational research, Burstein (1980) argued that existing statistical techniques were simply inadequate for estimating the effects of school on students. While Burstein's discussion focused on the research on school effects at the elementary and secondary level, the arguments and the methodological concerns he presents are also applicable to research focusing on the influence of college on students (Ethington, 1997). Burstein notes that the models used in school effects research had been single-level and based on the traditional linear model ordinary least squares (OLS) regression, which, he argues, does not adequately match the realities under investigation. Although researchers had acknowledged the hierarchical nature of the organization of schooling by gathering data on students, classes, and schools, the statistical model reflected only a single level. Burstein argued that this neglect of the hierarchical nature of the data gathered reflects the limitations of the existing statistical techniques at that time for the estimation of the linear models with nested structures rather than a conviction on the part of the researcher that the single-level statistical model was appropriate. There simply were no viable alternatives.

Historically, two common procedures have been used when analyzing hierarchical data. The first procedure is to disaggregate all higher order variables to the individual level, and the analysis is done at the individual level. The second procedure is to aggregate the individual level variables to the higher level and do the analysis at the higher level. However, a number of conceptual and technical difficulties such as aggregation bias, misestimated standard errors, and heterogeneity of regression have plagued these studies (Burnstein, 1980; Raudenbush & Bryk, 2002). These two procedures are known as the unit of analysis problem and have plagued researchers in their attempt to analyze hierarchical data.

Pascarella and Terenzini (1991) discuss the unit of analysis problem associated with the hierarchical nature of data in higher education and suggest that differences in the units of analysis used in studies examining similar phenomena may have contributed to the lack of consistency in findings in the influence of college on students. Pascarella and Terenzini reviewed studies that varied in the unit of analysis used and noted that after one statistically controls for the characteristics of students, the effects of attending different types of four-year institutions are both small and inconsistent. However, instead of concluding that different types of four-year institutions have essentially the same impact on student development, they offer an alternative explanation for the absence of institutional effects. They argue that student precollege characteristics are not independent of the institution attended, and that global college environment measures may have little impact on students given the subenvironments existing within institutions such as different majors and living arrangements that are more proximal to students' daily experiences. Essentially, Pascarella and Terenzini are acknowledging the multilevel nature of postsecondary institutions and its impact on research on college effects.

As Pascarella and Terenzini (1991) have documented, the unit of analysis issue has been a complex and controversial issue in the research on the influence of college.

The problem with disaggregating higher order variables to the individual level is the fact that if students are in the same institution then they have the same value on each institutional variable. Individuals in one group, whether it be individuals in the same major or individuals attending the same institution, are more similar than individuals in different groups. Thus students in different majors or institutions can be independent but students in the same majors or institutions share values on many more variables. Some of these variables are not observed, which means they vanish into the error term of the linear model, causing correlations between disturbances. The sharing of the same group is a likely cause of dependency among observations. To acknowledge the dependency of these individuals is important because it changes the error variance in traditional OLS regression. The error variance in traditional OLS regression represents the effect of all omitted variables and measurement error, under the assumption that these errors are unrelated. The degree of covariance in the error terms of individuals sharing the same institution or academic major is expressed in the intra-class correlation coefficient. OLS regression fails to capture the positive intra-class correlations that results from the interdependencies among students within the same institution, major, class, etc. These interdependencies are brought about by the common experiences of students within the same institution or because of the ways in which students were initially drawn to an institution and result in misestimated standard errors (Burstein, 1980).

In the second approach, student characteristics are aggregated over institutions and an institutional analysis is done. The procedure forces the researcher to assume that all individuals within the same institution are affected identically by the institutional level characteristics. Conceptually, this is an obvious error since institutions allocate financial resources differently through the institution. The main problem with this approach is the loss of the within-group information, which can usually account for up to 80 to 90% of the total variation (Burstein, 1980; Ethington, 1997; Hannan & Burstein, 1974; Kreft & de Leeuw, 1998). As a consequence, relations between aggregated variables are often much stronger and can be quite different from the relation between those at the individual level (Burstein, 1980; Draper, 1995; Kreft & de Leeuw, 1998). The aggregation approach is also problematic conceptually because student level characteristics change their meaning when aggregated; this is known as aggregation bias. Aggregation bias occurs when a variable takes on different meaning and therefore may have different effects at different levels of analysis.

For example, when analyzing workers in 12 different industries, Kreft, de Leeuw, and Aiken (1995) drew contradicting conclusions based on differing units of analysis. In their first analysis, executed at the level of the individual worker, the data showed a positive relationship between educational level and income: the higher the educational level, the higher the personal income. In the second analysis, executed at the level of the industry, the data showed a negative relationship between education and income: the higher the average educational level of an industry, the lower the average income of workers in that industry (colleges and universities are a good example of this). The industry-level analysis used aggregated measures, and these results illustrate that analyses executed at different levels of the hierarchy do not necessarily produce the same results. The fact that aggregate measures analyzed at the higher level of the hierarchy can produce results different from the original individual results has been well documented (Burstein, 1980; Robinson, 1950). An important conclusion can be drawn from these results. Since educational attainment had a positive effect on income if the unit of analysis is the individual and a negative effect on income if the unit of analysis is the industry, the logical conclusion is that the variable education measures different things depending on the unit of analysis. As Burstein (1980) argues, the issue is not that one unit of analysis is more appropriate than the other; rather the issue should be understood in light of the fact that different units of analysis are asking different questions of the data.

Therefore, a statistical analysis that can take into account the problems associated with the unit of analysis problem and model all levels of interests simultaneously is needed. Hierarchical linear models (HLM) solve the problems associated with the unit of analysis problem such as misestimated standard errors, heterogeneity of regression and aggregation bias by modeling all levels of interest simultaneously. Hierarchical linear modeling resolves the problem of misestimated standard errors by incorporating a unique random effect for each institution into the statistical model; moreover, the variability in these random effects is taken into account in estimating the standard errors (Raudenbush & Bryk, 2002). Until the advent of HLM, heterogeneity of regression had often been viewed as a methodological nuisance. However, the cause of heterogeneity of regression is often of substantive interest. HLMs enable a researcher to estimate a separate set of regression coefficients for each higher level organizational unit and then model variation among the higher level units in their sets of coefficients as multivariate outcomes to be explained by higher level factors (Raudenbush & Bryk, 2002). HLMs solve the problem of aggregation bias by modeling each level of the hierarchy with its own model.

Today, many higher education scholars are rushing to use this new, sophisticated analytic procedure (Smart, 2005). This rush seems to be based on the assumption that

HLM might yield substantively different findings than those from studies based on OLS regression analyses. With this in mind, the current study investigates the different conclusions that may be drawn depending upon the type of analysis chosen. I will focus on the three types of analyses discussed above. The first analysis will be an OLS regression with the student as the unit of analysis, the second analysis will be an OLS regression with the student level variables aggregated to the institutional level with the institution as the unit of analysis, and the third analysis will be a three-level hierarchical linear model with student characteristics modeled at Level 1, characteristics about the major modeled at Level 2 and characteristics of the institution modeled at Level 3.

Brief History of HLM

An early approach to dealing with the analytical problems associated with multilevel data was what had become known as the "slopes as outcomes" approach to regression (Burstein, 1980, Burstein, Linn, & Capell, 1978, Burstein & Miller, 1980). Burstein and colleagues estimated separate OLS regression equations for each school using only student-level predictors for a student-level outcome. They then used the regression coefficients from these equations as outcomes to be explained by school-level characteristics. This method was very appealing to researchers since it allowed for the relationships among student-level measures to be uniquely determined for each group using only within-group variability, and the variability predicted by the school-level measures represented between-school variability without the noise from the within-school variance affecting the between-school equations. However, the "slopes as outcomes" approach was incomplete. Since, the regression coefficients in OLS regression are estimated with considerable error, this limited the approach in detecting effects of between-group characteristics. The coefficient variance needs to be separated into its components in order to accurately test the group-level effects, and OLS regression is not able to analyze this complex variance-covariance structure (Ethington, 1997; Raudenbush & Bryk, 1986).

The term hierarchical linear model was first introduced by Lindley and Smith (1972) and Smith (1973) as part of their seminal study on Bayesian estimation of linear models. In their study, Lindley and Smith introduced a general framework for nested data with complex error structures. However, Lindley and Smith's contribution was not immediately able to be applied due to the fact that the model required estimation of covariance components in the presence of unbalanced data (Raudenbush & Bryk, 2002). No feasible estimation approach was available until Dempster, Laird, and Rubin (1977) developed the expectation-maximization (EM) algorithm. The EM algorithm provided an acceptable approach for the estimation of the covariance component. With the advent of advanced computer computations, the 1980s saw a resurgence in statistical theory and estimation procedures which led to a new class of statistical methods based on the hierarchical linear model.

As noted above, higher education data commonly have a nested structure, including, for example, students nested within academic majors. These academic majors are also nested within institutions. Further, the institutions may be nested within states, and even within countries. With hierarchical linear models, each of the levels in this structure is formally represented by its own submodel. These submodels express relationships among variables within a given level, and specify how variables at one level influence relations occurring at another (Raudenbush & Bryk, 2002).

Robustness Issues with OLS Regression

Multiple regression analysis is a versatile, all-purpose system for analyzing educational data (Cohen & Cohen, 1983). The concept of regression was first introduced by Sir Francis Galton (1886) while examining the relationship between fathers' and sons' heights. Galton observed that sons' heights do not tend toward their fathers' heights but instead regress toward the mean height of the population. Galton thus devised the first idea of regression and coupled with the method of least squares formulated by Carl Friedrich Gauss (1809), multiple regression analysis using ordinary least squares (OLS) procedures have become one of the most common statistical techniques for investigating and modeling relationships among variables.

As with all parametric statistics, the application of OLS regression and HLM analysis are based on certain assumptions. Understanding when violations of assumptions lead to serious biases and when they are of little consequence is essential to meaningful data analysis (Pedhazur, 1997). The assumptions underlying the application of ordinary least squares regression are (1) linearity; (2) no measurement error; (3) mean independence; (4) homoscedasticity; (5) uncorrelated errors; and (6) normally distributed errors. For the first assumption, linearity, it is assumed that the outcome can be expressed as a linear function of the independent variables and some random error term; it is further assumed that all of the relevant independent variables are included in the model. For the second assumption, no measurement error, it is assumed that each of the independent variables in the model is measured without error. The remaining assumptions are concerned with the errors. For the third assumption, mean independence, it is assumed that the mean of the error term is zero and that this value does not depend on the independent variables. The fourth assumption, homoscedasticity, states that the variance of the error term is the same across all levels of the independent variables. For the fifth assumption, uncorrelated errors, we assume that the values of the error term for any one observation are not influenced by the value of the error term for other observations. Finally, we assume that the overall distribution of the error term is normally distributed.

It has been demonstrated that regression analysis is generally robust against departures from assumptions with the exception of measurement errors and specification errors (Bohrnstedt & Carter, 1971; Pedhazur, 1997). Measurement errors in the dependent variable do not lead to biases in the estimation of the regression coefficient; however, they do lead to an increase in the standard error of estimate, thereby weakening the test of statistical significance. While measurement errors in the independent variables are more complex and the direction of bias may be in overestimation or underestimation (Pedhazur, 1997), attention is called to the importance of this issue where neglect of measurement errors in the independent variables can lead to misleading interpretations and conclusions. Specification errors refer to any errors committed in specifying the model to be tested. Such errors are omission of relevant variables from the equation, inclusion of irrelevant variables in the equations, and specifying that the regression is linear when a curvilinear relationship exists. Variable misspecification leaves its imprint on the error term and leads to violations of assumptions required for appropriate use OLS regression.

Robustness Issues with HLM

As Cohen and Cohen (1983) and Pedhazur (1997) show, the OLS regression model is highly versatile and this versatility carries over to multilevel regression analysis, or hierarchical linear modeling, which is essentially a multilevel extension of OLS regression. However, there is some evidence that one can come to different conclusions depending upon the type of analysis chosen, HLM or OLS regression (de Leeuw & Kreft, 1995; Hox, 1998; Webster, Mendro, Orsak, & Weerasinghe, 1996). A key difference between HLM and OLS regression is that HLM allows for the examination of varying effects between and within groups. The consequences of using a single level analysis, such as OLS regression, method on multilevel data are well-known: the parameter estimates are unbiased but inefficient and the standard errors are negatively biased, which results in spuriously significant effects (de Leeuw & Kreft 1986; Hox 1998, 2002; Maas & Hox, 2004). However, these biases are only in the presence of a large intra-class correlation.

In a three-level HLM, the following assumptions are made. First, as with OLS regression, it is assumed that the outcome at each level can be expressed as a linear function of the independent variables. The general three-level equations are as follows:

Level-1 Model: $Y_{ijk} = \pi_{0jk} + \pi_{1jk}\, a_{1ijk} + \pi_{2jk}\, a_{2ijk} + \ldots + \pi_{pjk}\, a_{pijk} + e_{ijk}$

Level-2 Model: $\pi_{pjk} = \beta_{p0k} + \beta_{p1k}\, X_{1jk} + \beta_{p2k}\, X_{2jk} + \ldots + \beta_{pqk}\, X_{qjk} + r_{pjk}$

Level-3 Model: $\beta_{pqk} = \gamma_{pq0} + \gamma_{pq1}\, W_{1k} + \gamma_{pq2}\, W_{2k} + \ldots + \gamma_{pqs}\, W_{sk} + u_{pqk}$

Second, the Level-l random effects (e_{ijk}) are assumed to be normally distributed with a mean of 0 and a constant variance, σ^2. Third, the Level-1 predictors (a_{pjk}) are independent of Level-1 random effect (e_{ijk}). Fourth, the set of Level-2 random effects (r_{pjk}) are assumed to be multivariate normally distributed each with a mean of 0, some variance τ_{pp}, and some covariance between r_{pjk} and $r_{p'jk}$ of $\tau_{pp'}$. Moreover, the random effects in Level-2 are assumed to be correlated. Fifth, the set of Level-2 predictors is independent of every r_{pjk}. Sixth, the Level-3 random effects (u_{pqk}) are assumed multivariate normally distributed with a mean of 0, some variance, and covariance among all pairs of elements. Seventh, the errors at all levels are independent. Finally, the predictors at each level are not correlated with the random effects at the other levels.

Simulation studies have been used to test the robustness of HLM. Since maximum likelihood estimation methods used in hierarchical linear modeling are asymptotic, sample sizes must be sufficiently large. An important issue in multilevel modeling is what constitutes a sufficient sample size for accurate estimation and the associated standard errors. The main problem is usually the sample size at the group level, because group-level sample size is always smaller than the individual-level sample size. Simulation studies have been used to address this problem. A review of the few simulation studies that have been carried out to date suggest that a large number of groups is generally more important than a large number of people per group (Kim, 1990; Maas & Hox 2004, 2005). The absolute minimum number of groups for accurate maximum likelihood estimation is debatable. Kreft and de Leeuw (1998) recommend that 30 groups is the absolute smallest acceptable number of groups for an HLM analysis. Maas and Hox (2004, 2005) recommend no less than 50 groups. They have shown that when sample sizes at Level 2 are less than 50 the standard errors of the Level-2 variance components are biased downward.

Maas and Hox (2004) also examined the assumption concerning the normality of the Level-2 residuals. When the Level-2 residuals are multivariate normally distributed, there is only a problem with the standard errors of the second level variances when the number of groups is less than 50 and group size is less than 30. When the Level-2 residuals are not normally distributed, only the standard errors for the random effects at Level 2 are highly inaccurate. With a large number of groups, the estimation of the fixed effects is unbiased even in the presence of nonnormally distributed residuals. In a later study, Maas and Hox (2005) confirmed these results showing that group sizes less than fifty leads to biased estimates of the second-level standard errors. The regression coefficients and the variance components were estimated without bias in all the simulated conditions. In addition, Maas and Hox (2004) recommend the following rule of thumb: "if one is only interested in the fixed effects of the model, ten groups can lead to good estimates. If one is also interested in contextual effects, 30 groups are needed. If one also wants correct estimates of the standard errors, at least 50 groups are needed" (p. 135).

In summary, all of these simulation studies generally concluded that with a small number of groups at the higher level the regression coefficients are estimated without bias while their standard errors tend to be biased downward; the variance components tend to be estimated too small with standard errors that tend to be biased downwards. In general, the effect of violation of the assumption of normally distributed residuals resembles the effect of small sample sizes: the regression coefficients and their standard errors show little or no bias, but variance components and their standard errors may be biased.

CHAPTER 2

METHODOLOGY

This study examined the influences on college seniors' perceived critical thinking ability three ways— (1) an ordinary least squares (OLS) regression with the student as the unit of analysis, (2) an OLS regression with the institution as the unit of analysis, and (3) a three-level hierarchical linear model (HLM) with student attributes modeled at Level 1, characteristics of the major modeled at Level 2, and characteristics of the institution modeled at Level 3— in order to illustrate the differing conclusions one may come to depending upon the type of analysis chosen. In all three analyses, students' perceived critical thinking ability was modeled as a function of student attributes, attributes of the student's major, and characteristics of the institutions the students attended. In order to better compare the results from the HLM analysis and the two OLS regression models, slope effects in the HLM analysis were constrained to be fixed. Thus, in the HLM analysis only the intercepts were allowed to vary across majors and institutions.

Sample

Data for this study were taken from the 2006 administration of the National Survey of Student Engagement (NSSE). The NSSE obtains information from a random sample of first-year and senior students about the nature of their undergraduate experiences and measures the extent to which students engage in effective educational practices (Kuh, 2001). In the 2006 NSSE administration, 1,139,412 first-year and senior students from 557 institutions in the United States and Canada were eligible to participate. From this population of students, NSSE randomly sampled an equal number of first-year and senior students at each institution. This cohort of 752,675 randomly selected students compromised the 2006 NSSE sampling frame. Of those sampled, 259,679 students responded yielding a response rate of 35%. The institutions that participated in NSSE 2006 were very similar to the national profile of all baccalaureate degree-granting institutions in the United States in terms of sector, geographic region, and urban-rural locale.

The sample used in this study consists of senior students who completed the NSSE survey in 2006. Only students who had begun college at their current institution were selected for the sample. The restriction to include only students who had begun at their current institution was made in order to examine institutional effects. Students that had transferred to their current institution may not have had time to gauge important contributions of the institution. Next, institutions and majors with less than 30 students were omitted from the sample. This restriction was made in reference to Maas and Hox's (2004, 2005) recommendation on appropriate sample sizes for HLM analyses. The final sample used in this study consists of 56,276 senior students in 58 majors from 405 U.S. institutions that started college at their current institution and who had complete data on the variables described below.

The Model

For the purpose of this study, the dependent variable is self-perceived growth in critical thinking skills and is perceived to be a function of student attributes, the influence of the student's major, and attributes of the institution they attend. The variables chosen for this study to operationalize student attributes, college major attributes, and institutional attributes were selected from Pascarella and Terenzini's (2005) review of the literature of college effects on students. The student characteristics hypothesized to impact critical thinking are three scales measuring course emphasis on higher-order thinking skills, students' level of academic effort, and student-faculty interaction. Two of Biglan's (1973a, 1973b) three dimensions, hard vs. soft and pure vs. applied, were used to measure influences from the major. The characteristics of the institution hypothesized to impact critical thinking are measures of students' perceptions of supportive campus environment, the selectivity of institution, the extent of graduate emphasis, and the residential character of the institution.

This study estimated the effects of the above variables in three different ways. First, ordinary least squares regression with the student as the unit of analysis. Second, data were aggregated at the institution level and an ordinary least squares regression with the institution as the unit of analysis was estimated. Third, a three-level hierarchical linear model with student attributes at the first-level, attributes of the major at the second-level, and institutional attributes at the third-level. Thus, three statistical models are driving this study.

Variables

The variables used in this study were constructed from items included in the 2006 administration of the NSSE survey, Biglan's (1973a, 1973b) classification of academic disciplines, the 2005 Carnegie advanced classification, and Barron's ratings of institutional selectivity. The dependent variable used in the analyses was a scale representing student's perceived critical thinking ability (CT). The NSSE survey asked students questions regarding the extent to which their experiences at their current institution contributed to their knowledge, skills, and personal development in thinking critically and analytically, analyzing quantitative problems, and solving complex real-world problems. The alpha reliability coefficient for perceived critical thinking ability was 0.79. Appendix A provides a complete list of the items comprising each variable along with the coding and construction procedures.

The student characteristics hypothesized to impact critical thinking are three scales representing course emphasis on higher-order thinking skills, academic effort, and student-faculty interaction. The six items comprising the course emphasis on higher-order thinking skills scale (HOT) ask students the extent to which their coursework emphasized analyzing and synthesizing ideas, making judgments, and applying theories. The 11 items comprising the level of academic effort scale (AE) ask students questions related to course rigor and preparation. The five items comprising the student-faculty interaction scale (SFI) ask student about discussions and interactions with faculty members. Alpha reliability coefficients for these scales are 0.80, 0.67, and 0.77, respectively. The selection of items for these scales was taken from Pike, Kuh, and McCormick (2008) except for one item in the student-faculty interaction scale. The item that was omitted asked students how often they had received prompt feedback from faculty on their academic performance. This question was omitted because of the vagueness in the language regarding *prompt* feedback. In their study, Pike et al. found alpha reliability coefficients similar to the ones found in this study.

The major characteristics hypothesized to impact critical thinking are Biglan's (1973a, 1973b) hard versus soft dimension and pure versus applied dimension. The hard versus soft dimension (HARD) reflects the degree to which an academic discipline possesses a clearly delineated paradigm. The pure versus applied dimension (PURE) reflects the academic discipline's concern with practical application. Each major was classified as either hard or soft and either pure or applied. Thus, for example mathematics is classified as both "Hard" and "Pure" whereas finance is classified as both "Soft" and "Applied." The hard versus soft dimension is coded 0 for soft disciplines and 1 for hard disciplines. The pure versus applied dimension is coded 0 for applied disciplines and 1 for pure disciplines. Appendix B lists all the majors and their Biglan classification.

The institutional characteristics hypothesized to impact students' perceived critical thinking are measures of the supportive campus environment, the selectivity of institution, the extent of graduate emphasis, and the residential character of the institution. The six items comprising the supportive campus environment scale (SCE) ask students questions about their institutions commitment to their academic and social success and their relationships with other students, faculty members, and administrative personnel. Since the supportive campus environment scale is considered an institutional characteristic it was aggregated for each institution. The supportive campus environment scale represents a characteristic of the normative institutional environment and is the average perception of the supportive environment of the institution. Alpha reliability coefficient for the supportive campus environment scale is 0.78. The selection of items for the supportive campus environment scale was also based on Pike et al. (2008). Furthermore, the alpha reliability coefficient computed by Pike et al. was similar to the one found in this study.

In addition, two of the 2005 Carnegie advanced classifications were used. The first classification, graduate coexistence, measures the extent to which an institution awards graduate degrees in the same fields in which they award undergraduate degrees.

The values in the graduate coexistence variable were merged into three categories: no graduate coexistence, some graduate coexistence, and high graduate coexistence. Then, two dummy variables were created. One dummy variable (SG) was coded 1 for some graduate coexistence and 0 otherwise. The other dummy variable (HG) was coded 1 for high graduate coexistence and 0 otherwise. Thus, no graduate coexistence was the comparison variable. The next classification measures the institutions' residential character. The values in the residential character variable were merged into three categories: primarily commuter, primarily residential, and highly residential campuses. Then, two dummy variables were created. One dummy variable (PC) was coded 1 for primarily commuter and 0 otherwise. The other dummy variable (PR) was coded 1 for primarily residential and 0 otherwise. Thus, highly residential was the comparison variable. The final institutional characteristic used in this study is the 2005 Barron's ratings of institutional selectivity (BAR). This index has 11 categories ranging from "noncompetitive" to "most competitive."

Data Analysis

This study estimated the effects of the above variables in three different ways. The first analysis was an OLS regression with the student as the unit of analysis. The second analysis was an OLS regression with the institution as the unit of analysis. The third analysis was a three-level HLM analysis with student characteristics modeled at Level 1, characteristics of the academic discipline modeled at Level 2, and characteristics of the institutions modeled at Level 3. Prior to the estimation of the two OLS regression models and the HLM model, exploratory analyses were conducted testing the assumptions underlying each of the analyses. Normal probability and residual plots indicated that the OLS regression assumptions of normality and heterogeneity of variance were satisfied. Moreover, residual statistics were checked for any potential outliers and influential data points. Variance inflation factors (VIF) were calculated and results indicated that multicollinearity was not an issue in the data.

For the three-level HLM analysis, model assumptions were checked by comparing the results of the model-based fixed effects with the results of the fixed effects with robust standard errors. Since the number of Level-3 units is relatively large, the model-based fixed effects can be compared to the fixed effects with robust standard errors. If the model-based fixed effects and the fixed effects with robust standard errors differ substantially, it suggests problems with normality, homosecdasticity, or linearity (Raudenbush & Bryk, 2002). In this analysis, model-based fixed effects and fixed effects with robust standard errors were similar suggesting no severe violations of the assumptions underlying the application of hierarchical linear modeling.

The first analysis was an OLS regression with the student as the unit of analysis. The equation estimated was

$$CT = \beta_0 + \beta_1 (HOT) + \beta_2 (AE) + \beta_3 (SFI) + \beta_4 (HARD) + \beta_5 (PURE) + \beta_6 (SCE) + \beta_8 (SG) + \beta_9 (HG) + \beta_{10} (PC) + \beta_{11} (PR) + \beta_{12} (BAR) + \varepsilon$$

In this analysis all variables were measured at the student-level except the supportive campus environment scale, which was aggregated to the institution-level. The second analysis was an OLS regression with the institution as the unit of analysis. The equation estimated was the same as the previous equation except that all variables were aggregated to the institution-level. Appendix C and D present the means, standard deviations, and correlations used for estimating the two OLS regression models.

The third analysis was the three-level HLM model. Since the HLM program requires three raw data files as input and does not have the capability for general exploration and manipulation of data, all preliminary analyses checking, cleaning, exploring the data, recoding and transforming variables were conducted using SPSS. Three raw data files were created. The first dataset contained information on the individual college students (the Level 1 file) while the second dataset contained information on the characteristics of the students' academic majors (Level 2 file), and the third dataset contained information on the characteristics of the institutions that those students attend (the Level 3 file). Each student's record contained a common Level-2 ID and Level-3 ID that links the student to a particular Level-2 major and Level-3 institution, respectively.

The HLM analysis was conducted in four phases. The first phase begins by estimating a model that has no Level-1, Level-2, or Level-3 predictors. The purpose of estimating a model with no predictors was to represent how the variation in students' perceived critical thinking ability was allocated across the three different levels (student, major, and institution). Raudenbush and Bryk (2002) refer to this model as the fully unconditional model since there are neither student-level predictors used at Level1 or any major or institutional characteristics as predictors at Level 2 or Level 3. The Level-1 equation is

$$\mathrm{CT}_{ijk} = \pi_{0jk} + \mathrm{e}_{ijk}$$

where

CT_{ijk} is the perceived critical thinking skills of student i in major j and institution k;

π_{0jk} is the mean critical thinking score of major *j* in institution *k*;

e_{ijk} is the random "student effect," that is, the deviation of student *ijk's* score from the major mean. These effects are assumed normally distributed with a mean of 0 and variance σ^2.

The indices *i*, *j*, and *k* denote students, majors, and institutions where there are

$i = 1, 2, \ldots, n_{jk}$ students within major *j* in institution *k*;

$j = 1, 2, \ldots, J_k$ majors within institution *k*; and

$k = 1, 2, \ldots,$ K schools.

Each student's critical thinking skills are characterized as a function of his or her major average critical thinking score, π_{0jk}, and a random effect, e_{ijk}. The variance of the random effect is denoted σ^2 and represents the pooled within-major variance (or variance among students).

At the second level, each major mean, π_{0jk}, is viewed as an outcome varying randomly around some school mean. The Level-2 equation is

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$

where

β_{00k} is the mean critical thinking score in institution *k*;

r_{0jk} is the random "major effect," that is, the deviation of major *jk's* mean from the institution mean. These effects are assumed normally distributed with a mean of 0 and variance τ_π.

Within each of the K institutions, the variability among majors is assumed the same.

The Level-3 model represents the variability among institutions. The institution means, β_{00k}, are viewed as varying randomly around the grand mean. The Level-3 equation is

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

where

γ_{000} is the grand mean;

u_{00k} is the random "institution effect," that is, the deviation of institution *k's* mean from the grand mean. These effects are assumed normally distributed with a mean of 0 and variance τ_β.

The fully unconditional three-level model partitions the total variability in critical thinking skills into its three components: among students within majors, σ^2; among majors within institutions, τ_π; and among institutions, τ_β. It also allows for the estimation of the proportion of variation that is within majors, among majors within institutions, and among institutions. That is,

$\sigma^2 / (\sigma^2 + \tau_\pi + \tau_\beta)$ is the proportion of variance within majors (student-level variance pooled within majors);

$\tau_\pi / (\sigma^2 + \tau_\pi + \tau_\beta)$ is the proportion of variance among majors within institutions (major-level variance pooled among majors within institutions); and

$\tau_\beta / (\sigma^2 + \tau_\pi + \tau_\beta)$ is the proportion of variance among intuitions (institution-level variance across institutions).

In the second phase, a full Level-1 model was estimated using the students' characteristics to predict students' perceived critical thinking ability. Within each major, students' perceived critical thinking ability was modeled as a function of student-level predictors plus a random student-level error. In this model, only the intercept was allowed to vary across majors; slope effects were constrained to be fixed across majors and institutions. Considering students' perceived critical thinking ability to be a function of course emphasis on higher-order thinking skills, academic effort, and student-faculty interaction, the following equation was estimated for each major:

$$CT_{ijk} = \pi_{0jk} + \pi_{1jk}(HOT_{ijk}) + \pi_{2jk}(AE_{ijk}) + \pi_{3jk}(SFI_{ijk}) + e_{ijk}$$

Each of the student-level predictors were centered about the major mean, and thus,

π_{0jk} is the average across majors;

π_{1jk} is the effect of course emphasis on higher-order thinking skills on critical thinking;

π_{2jk} is the effect of academic effort on critical thinking;

π_{3jk} is the effect of student-faculty interaction on critical thinking; and

e_{ijk} is the student-level random effect that represents the deviation of student *ijk's* score from the predicted score based on the student-level model. These residual student effects are assumed normally distributed with a mean of 0 and variance σ^2.

The third phase tested whether the effects of the intercept, π_{0jk}, varied across majors. Since significant variability was found, π_{0jk} was modeled as a function of the two Level-2 variables. In order to better compare the results of the HLM analysis with the two OLS regression analyses, the slope effects were fixed to equal the average across majors; only the intercept was allowed to vary. Thus, the following equation was estimated for each major:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(HARD) + \beta_{02k}(PURE) + r_{0jk}$$

where

β_{00k} is the intercept for institution *k* in modeling the major effect π_{0jk};

β_{01k} is the corresponding coefficient that represents the direction and strength of association between major characteristic (HARD) and π_{0jk};

β_{02k} is the corresponding coefficient that represents the direction and strength of association between major characteristic (PURE) and π_{0jk}; and

r_{0jk} is a Level-2 random effect that represents the deviation of major *jk's* Level-1 coefficient, π_{0jk}, from its predicted value based on the major-level model.

The random effects in these equations are assumed to be correlated. Formally, it is assumed that the set of r_{0jk} are multivariate normally distributed each with a mean of 0, some variance τ_{pp}, and some covariance between elements r_{0jk} and $r_{0'jk}$ of $\tau_{pp'}$. These variances and covariances are collected in a matrix labeled $\mathbf{T}_{\pi}$ whose dimensionality depends on the number of Level-1 coefficients specified as random.

The final phase tested whether the effects of the intercept, β_{00k}, varied across institutions. Since significant variability was found, β_{00k} was modeled as a function of the four Level-3 variables. The slopes were fixed to equal the average across institutions; moreover, the supportive campus environment scale and Barron's rating of institutional selectivity variable were entered into the model centered around the grand mean. Thus, the following equation was estimated for each institution:

$$\beta_{00k} = \gamma_{000} + \gamma_{001}\,(SCE) + \gamma_{002}\,(SG) + \gamma_{003}\,(HG) + \gamma_{004}\,(PC) + \gamma_{005}\,(PR) + \gamma_{006}\,(BAR) + u_{00k}$$

where

γ_{000} is the intercept term in the institution-level model for β_{00k};

$\gamma_{00S,\ where\ S\ =\ 1,\ldots,6}$ is the corresponding Level-3 coefficient that represents the direction and strength of association between the institution characteristic and β_{00k}; and

u_{00k} is a Level-3 random effect that represents the deviation of school *k's* coefficient, β_{00k}, from its predicated value based on the institution-level model.

The residuals from these equations are assumed multivariate normally distributed. Each is assumed to have a mean of zero, some variance, and covariance among all pairs of elements. Here too, the variances and covariances are collected in a matrix, $\mathbf{T}_\beta$. The dimensionality of $\mathbf{T}_\beta$ depends on the number of Level-2 coefficients that are specified as random. All other β coefficients will be viewed as fixed thus their residuals are assumed to be zero.

CHAPTER 3

RESULTS

Student-level OLS Regression

In the first analysis, ordinary least squares regression was used to examine the influences on students' perceived critical thinking ability. The regression results indicated that the set of independent variables explained 31.9% of variance in critical thinking (F (11, 56264) = 2396.09, $p < .001$). Regression results are given in Table 1. In the presence of the other variables in the model, the variables which had a significant, unique relationship with students' perceived critical thinking ability were course emphasis on higher order thinking skills (b = 0.394), academic effort (b = 0.144), student-faculty interaction (b = 0.110), hard vs. soft dimension (b = 3.028), pure vs. applied dimension (b = -2.216), perceptions of supportive campus environment (b = 0.075), high graduate coexistence (b = 0.759), primarily commuter institutions (b = 1.291), and primarily residential institutions (b = 0.760). Only two variables did not have a significant impact on perceived gains in critical thinking: some graduate coexistence, and Barron's ratings of institutional selectivity.

Institution-level Regression

The second analysis was an OLS regression with the institution as the unit of analysis. In this analysis, all variables were aggregated to the institution-level. The regression results indicated that the set of independent variables explained 66.2% of variance in critical thinking (F (11, 393) = 69.99, $p < .001$). Regression results are given in Table 2. In the presence of the other variables in the model, the variables which had a significant, unique relationship with perceived critical thinking ability were course

Table 1

Student-level OLS Regression Results

Independent Variables	*b*	*S.E.*	*β*	*t*
Higher Order Thinking Skills (HOT)	0.394	0.004	0.394	91.890*
Academic Effort (AE)	0.144	0.004	0.144	35.73*
Student-Faculty Interaction (SFI)	0.110	0.004	0.111	26.65*
HARD	3.028	0.082	0.131	36.84*
PURE	-2.216	0.072	-0.111	-30.70*
Supportive Campus Environment (SCE)	0.075	0.005	0.065	15.54*
Some graduate coexistence (SG)	-0.024	0.102	-0.001	-0.23
High graduate coexistence (HG)	0.759	0.129	0.034	5.89*
Primarily commuter (PC)	1.291	0.119	0.056	10.84*
Primarily residential (PR)	0.760	0.096	0.036	7.92*
Barron's ratings of institutional selectivity (BAR)	0.069	0.037	0.008	1.87
R-square = 0.319				

*$p < .001$.

Table 2

Institution-level OLS Regression Results

Independent Variables	*b*	*S.E.*	*β*	*t*
Higher Order Thinking Skills (HOT)	0.485	0.047	0.499	10.34*
Academic Effort (AE)	0.192	0.040	0.207	4.85*
Student-Faculty Interaction (SFI)	0.026	0.041	0.034	0.64
HARD	4.047	0.412	0.322	9.83*
PURE	-2.726	0.465	-0.248	-5.86*
Supportive Campus Environment (SCE)	0.075	0.009	0.322	7.96*
Some graduate coexistence (SG)	-0.057	0.177	-0.013	-0.32
High graduate coexistence (HG)	0.303	0.252	0.054	1.20
Primarily commuter (PC)	1.328	0.217	0.271	6.13*
Primarily residential (PR)	0.674	0.170	0.149	3.97*
Barron's ratings of institutional selectivity (BAR)	0.027	0.073	0.014	0.37
R-square = 0.662				

*$p < .001$.

emphasis on higher order thinking skills (b = 0.485), academic effort (b = 0.192), hard vs. soft dimension (b = 4.047), pure vs. applied dimension (b = -2.726), perceptions of supportive campus environment (b = 0.075), primarily commuter institutions (b = 1.328), and primarily residential institutions (b = 0.674). Four variables did not have a significant influence on students' aggregated perceived critical thinking ability: student-faculty interaction, some graduate coexistence, high graduate coexistence, and Barron's ratings of institutional selectivity.

Hierarchical Linear Modeling Analysis

The third analysis was a three-level hierarchical linear model with student attributes modeled at Level 1, attributes of the major modeled at Level 2, and institutional characteristics modeled at Level 3. The three-level hierarchical linear model was analyzed using HLM 5.05 (Raudenbush, Bryk, & Congdon, 2000). The first step in the hierarchical linear modeling process involved determining how the variation in critical thinking was distributed among the three different levels: student, major, and institution. This was accomplished by estimating the fully unconditional model with no predictors at any of the three levels. Table 3 gives the results of the estimation of the fully unconditional model. The estimation of the grand mean of critical thinking across all majors within all institutions (the fixed effect) is 49.97. Decomposing the total variability in critical thinking into its' three components the estimates for the variability among students within majors (σ^2), among majors within institutions (τ_π), and among institutions (π_β) are 91.586, 5.075, and 2.510, respectively. Using these parameter estimates the intra-class correlations can be calculated (see p. 23 for formulas to compute intra-class correlations). In this case, the proportion of variance among students within majors was

Table 3

HLM Estimation of Unconditional Model

Fixed Effects	*Coefficient*	*S.E.*	*t-ratio*
γ_{000}: average student critical thinking score	49. 97	0.097	514.129*
Random Effects	*Variance*	*DF*	*Chi-square*
σ^2: variance among student within majors	91.586		
τ_{π}: variance among majors within institutions	5.075	9036	11965.90*
τ_{β}: variance among institutions	2.510	404	1298.32*
Intra-class Correlations	*Coefficient*		
Proportion of variance among students	92.4%		
Proportion of variance among majors	5.1%		
Proportion of variances among institutions	2.5%		

*$p < .001$.

92.4%, the proportion of variance among majors within institutions was 5.1%, and the proportion of variances among institutions was 2.5%. Chi-square test indicate that critical thinking scores vary significantly among majors within institutions (χ^2 (9036) = 11965.90, $p < .001$) and vary significantly among institutions (χ^2 (404) = 1298.32, $p < .001$). This variability will subsequently be modeled by using characteristics of the majors to predict π_{0jk} (student-level intercept) and institutional measures to predict β_{00k} (major-level intercept).

In the second step, a full Level-1 model was estimated using the students' characteristics to predict students' perceived critical thinking ability. In this analysis, slope effects of the student-level variables were fixed to be equal to average across majors within institutions. All three student-level variables were centered around their respective group means, so that the intercept, π_{0jk}, would represent the average critical thinking score across majors within institutions. This step was performed in order to estimate the proportion of variance in critical thinking ability among students within majors explained by the addition of the student-level predictors. The addition of the student-level variables (course emphasis on higher order thinking skills, academic effort, and student-faculty interaction) explained 27.7% of student-level variance. Additionally, the chi-square test revealed significant variation in the intercept, π_{0jk}, across majors (χ^2 (9036) = 15721.68, $p < .001$). Table 4 gives the results of the HLM estimation of random effects for the Level-1 model.

Since the intercept, π_{0jk}, varies across majors, it was modeled as a function of the Level-2 variables. In this analysis, student-level slopes were fixed to equal the average across majors within institutions and the major-level slopes were fixed to equal the

Table 4

HLM Estimation of Random Effects with Student-level Predictors (Level-1 Model)

Random Effects	*Variance*	*DF*	*Chi-square*
σ^2: variance among student within majors	66.193		
τ_π: variance among majors within institutions	9.500	9036	15721.68*
τ_β: variance among institutions	2.549	404	1266.26*

*p < .001.

average across institutions. Both major-level variables were entered into the model uncentered, so that the intercept, β_{00k}, represents the average critical thinking score across majors within institutions. The major-level variables were entered uncentered because they were dichotomies unlike the continuous variables at the student-level. The addition of the major-level variables (hard vs. soft dimension and pure vs. applied dimension) explained 21.88% of the variance in the student-level intercept, π_{0jk}. In other words, the Biglan (1973a, 1973b) variables, hard versus soft dimension and the pure versus applied dimension, explained 21.88% of variance between majors. Chi-square test indicate that the remaining unexplained variability is still significant (χ^2 (9034) = 14630.10, p < .001) indicating that there still are significant differences among the mean critical thinking levels of majors not explained by the two Biglan variables. Additionally, the chi-square test revealed significant variation in the major-level intercept, β_{00k} (χ^2 (404) = 1379.57, p

< .001) across institutions. Table 5 gives the results of the HLM estimation of random effects for the Level-2 model.

Table 5

HLM Estimation of Random Effects with Student-level and Major-level Predictors (Level-2 Model)

Random Effects	*Variance*	*DF*	*Chi-square*
σ^2: variance among student within majors	66.352		
τ_π: variance among majors within institutions	7.422	9034	14630.10*
τ_β: variance among institutions	2.606	404	1379.57*

*$p < .001$.

In the final step, the full HLM analysis was modeled. Since the intercept, β_{00k}, varies across institutions, it was modeled as a function of the Level-3 variables. In this analysis, student-level slopes were fixed to equal the average across majors within institutions and major-level slopes were fixed to equal the average across institutions. In this analysis, the supportive campus environment scale and Barron's ratings of institutional selectivity were entered into the model centered around their respective grand means and the other variables were entered into the model uncentered, so that the intercept, γ_{000}, represents the average critical thinking score across institutions. Results for the random effects are found in Table 6, and results for the fixed effects are found in Table 7. The addition of the institution-level variables explained 57.66% of the variance in major-level intercept, β_{00k}. In other words, the addition of the institutional characteristics explained 57.66% of the variance between institutions. Chi-square test indicate that the remaining unexplained variability is still significant (χ^2 (398) = 858.17, $p < .001$) indicating that there still are significant differences among the mean critical thinking levels of institutions not explained by the six institutional characteristics.

Table 6

HLM Estimation of Random Effects with Predictors Modeled at All Three Levels (Level-3 Model)

Random Effects	*Variance*	*DF*	*Chi-square*
σ^2: variance among student within majors	66.370		
τ_π: variance among majors within institutions	7.415	9034	15445.10*
τ_β: variance among institutions	1.103	398	858.17*

**p* < .001.

The final estimation of the fixed effects was as follows. All three student-level variables had a significant impact on students' perceived critical thinking ability. The estimated effect for course emphasis on higher order thinking skills was 0.387; the estimated effect for academic effort was 0.136, and the estimated effect of student-faculty interaction was 0.125. Both major-level variables had a significant impact on average critical thinking across major within institution. The estimated effect for the hard vs. soft dimension was 2.655, and the estimated effect for the pure vs. applied dimension was -2.111. Four of the six institution-level variables had a significant effect on the average critical thinking across institutions. The estimated effect for supportive campus environment was 0.140; the estimated effect for primarily residential was 0.584; the estimated effect for primarily commuter was 1.310, and the estimated effect for Barron's ratings of institutional selectivity was 0.401. The estimated effect for the intercept, γ_{000}, was 50.211.

Table 7
HLM Estimation of Fixed Effects with Predictors Modeled at All Three Levels

Fixed Effects	*Coefficient*	*S.E.*	*t-ratio*
Intercept	50.211	0.187	268.65**
Level 1: Effects on student critical thinking			
Higher Order Thinking Skills (HOT)	0.387	0.005	80.89**
Academic Effort (AE)	0.136	0.005	30.20**
Student-Faculty Interaction (SFI)	0.125	0.005	26.84**
Level 2: Effects on average critical thinking across majors within institutions			
HARD	2.655	0.121	21.99**
PURE	-2.111	0.106	-19.99**
Level 3: Effects on average critical thinking across institutions			
Supportive Campus Environment (SCE)	0.140	0.010	13.87**
Some Graduate Coexistence (SG)	-0.470	0.208	-2.27
High Graduate Coexistence (HG)	-0.054	0.284	-0.19
Primarily Commuter (PC)	1.310	0.249	5.27**
Primarily Residential (PR)	0.584	0.204	2.86*
Barron's ratings of institutions selectivity (BAR)	0.401	0.080	5.03**

**p* < .01. ***p* < .001.

CHAPTER 4

DISCUSSION

Examining a sample of college seniors who took part in the 2006 administration of National Survey of Student Engagement, this study investigated the influences on seniors' perceived critical thinking ability three ways in order to illustrate the differing conclusions one may come to depending upon the type of analysis chosen. The first analysis was an ordinary least squares (OLS) regression with the student as the unit of analysis. The second analysis was an OLS regression with the institution as the unit of analysis. The third analysis was a three-level hierarchical linear model (HLM) with student attributes modeled at Level 1, characteristics of the major modeled at Level 2, and characteristics of the institution modeled at Level 3. The differences in results are noted and discussed in terms of substantive differences in the conclusions drawn from the analyses depending on the type of methodology is used. Furthermore, a comparison of coefficient estimates and standard errors are discussed and compared across analyses along with issues regarding sample sizes.

OLS Regression with the Student as the Unit of Analysis

In the field of higher education, researchers studying college effects on students generally use the student as the unit of analysis. Often times, these studies contain mixed forms of data. Researchers acknowledge the importance of the hierarchical nature of the organization of postsecondary education, which is why they typically collect information about students and characteristics of the institutions they attend. Thus, if we perform the analysis as the majority of higher education researchers would, we would come to the following conclusions regarding the influences on students' perceived critical thinking ability.

Results from the OLS regression with the student as the unit of analysis (Table 1) indicate that the set of independent variables explain 31.9% of the variance in students' perceived critical thinking ability. From the results, we see that all three student attributes have a significant impact on students' perceived critical thinking ability; furthermore, course emphasis on higher order thinking skills has the strongest relationship to students' perceived critical thinking ability (b = 0.394, β = 0.394). Thus, coursework that emphasizes analyzing, synthesizing, and making judgments about ideas and information, applying theories or concepts to new situation, integrating ideas from various sources of information, and putting together ideas or concepts from different courses when completing assignments, leads to higher perceptions of critical thinking ability than coursework that does not emphasize these types of learning. The second strongest relationship was students' level of academic effort (b = 0.144, β = 0.144). The greater investment of time and effort students put into their academic work, the greater the perceived gains in critical thinking. Student-faculty interaction (b = 0.110, β = 0.111) is also shown to have a positive, significant relationship to students' perceived critical thinking ability indicating that the more time and effort students spend interacting with faculty members the greater the perceived gains in critical thinking.

Both major characteristics have a significant influence on students' perceived critical thinking ability. Students majoring in hard fields (b = 3.028, β = 0.131), i.e., academic disciplines that have a commonly agreed upon set of problems for study and accepted methods for exploring these problems, tend to perceive greater critical thinking ability than students majoring in soft fields. Furthermore, students majoring in applied fields (b = -2.216, β = -0.111), i.e., an academic discipline that is concerned with the practical application of its subject material, tend to perceive greater critical thinking ability than students majoring in pure fields.

The following institutional characteristics have a significant influence on students' perceived critical thinking ability: supportive campus environment (b = 0.075, β = 0.065), high graduate coexistence (b = 0.759, β = 0.034), primarily commuter institutions (b = 1.291, β = 0.056), and primarily residential institutions (b = 0.760, β = 0.036). The greater students perceived the campus as a supportive and friendly place the greater the perceived gains in critical thinking. Additionally, students attending institutions with a high graduate coexistence perceive greater gains in critical thinking than student who attend institutions with no graduate coexistence. Interestingly, students that attended institutions that were not highly residential institutions, in other words, institutions that were primarily commuter or primarily residential, perceive greater critical thinking ability than students that attended intuitions that were highly residential. Selectivity, as measured by Barron's ratings of institutional selectivity, was not found to have a significant relationship with students' perceived critical thinking ability.

Although the supportive campus environment scale, high graduate coexistence and residential character variables have a significant influence on students' perceived critical thinking ability, the statistical significance could be due to the large sample size (n = 57,276) used in the analysis. While this large sample size was not required for the OLS regression model, it was needed in order to meet the appropriate sample size requirements recommended for HLM analyses by Maas and Hox (2004, 2005) and Kreft and de Leeuw (1998). In light of the large sample size, Pedhazur (1997) argues that standardized regression coefficients smaller than 0.05, regardless of probability level, are substantively not worth interpreting. Pedhazur reasons that when sample size is relatively large, even substantively meaningless regression coefficients may be statistically significant. Consequently, researchers should use a criterion of meaningfulness, specific to the area of study, when interpreting significant regression coefficients. Given Pedhazur's argument and the reality of the large sample size, it is reasonable to assume that, according to this analysis, the effects of institutional characteristics are minimal at best.

A Comparison of the Two OLS Regression Models

An important question when investigating the influence of college on students is the appropriate unit of analysis. Generally, higher education researchers use the student as the unit of analysis when studying college effects on students. An alternative approach, prior to more advanced statistical techniques, was to aggregate the student-level data to the institution-level and perform the analysis on the institution. If this approach were taken to analyze the data, we would have come to the following conclusions regarding the influences on the average student's perceived critical thinking ability.

Results from the OLS regression with the institution as the unit of analysis (Table 2) indicate that the set of independent variables explain 66.2% of the variance in the average student's perceived critical thinking ability. In the institution OLS regression model, two of the three student characteristics have a significant influence on the average students' perceived critical thinking ability. Just as in the previous analysis, course emphasis on higher order thinking skills (b = 0.485, β = 0.499) has the greatest impact on critical thinking ability. In addition, students' level of academic effort (b = 0.192, β = 0.207) also has a significant impact on critical thinking. Different from the student level analysis, student-faculty interaction was not shown to significantly impact the average students' critical thinking ability.

Again, both major characteristics have a significant influence on the average students' perceived critical thinking ability. Students majoring in hard fields (b = 4.047, β = 0.322), tend to perceive greater critical thinking ability than students majoring in soft fields. Furthermore, students majoring in applied fields (b = -2.726, β = -0.248) tend to perceive greater critical thinking ability than students majoring in pure fields. The institutional characteristics that have a significant influence on the average student's perceived critical thinking ability are the same as in the OLS regression analysis with the student as the unit of analysis except for the effect of high graduate coexistence. Institutions where students perceive the campus as a supportive and friendly place tend to have average student bodies that also perceive greater critical thinking ability (b = 0.075, β = 0.322). Moreover, the average student body at institutions that are primarily commuter (b = 1.328 β = 0.271) or primarily residential (b = 0.674, β = 0.149) perceive greater gains in critical thinking than the average student body at institutions that are highly residential. In the institution OLS regression model, the two graduate coexistence variables and selectivity do not have a significant relationship with the average student's perceived critical thinking ability.

It was noted that in the student OLS regression analysis, the institutional variables had marginal effects. Conversely, in the institution OLS regression analysis all significant standardized coefficients were very strong. Moreover, the standardized coefficients for all significant variables were larger in the institution OLS regression model than in the student OLS regression model. The stronger relationships found in the institution OLS regression analysis were expected given the citations in the literature (Burstein, 1980; Draper, 1995; Kreft & de Leeuw, 1998) that relations between aggregated variables are often times stronger. Another instance where aggregate data tends to be stronger is in the estimation of the variance explained. The estimate of variance explained in the institution analysis appears much larger than the amount of variance explained in the student model. Given the stronger relationships in the institution OLS regression model, we would expect to see a larger proportion of variance explained. The variance explained in the institution analysis appears larger because we ignore the individual variability and only have the variability that is between institutions, which is a much smaller proportion. This will become more apparent when we look at the variance decomposition in the HLM analysis. Finally, the results of the institution OLS regression analysis are not as affected by sample size (n = 407) as they are in the student OLS regression model. Thus, the statistical significance of the variables in the institution OLS regression is much more reliable.

Appropriateness of Hierarchical Liner Modeling

In the past, the unit of analysis problem plagued higher education researchers in their attempt to study college effects on students. The two most common procedures to address the unit of analysis problem is to either disaggregate all higher order variables to the lower level and perform the analysis at the lower level, as was done in the OLS regression analysis with the student as the unit of analysis, or aggregate all lower level variables to the higher level and perform the analysis at the higher level, as was done in the OLS regression analysis with the institution as the unit of analysis. However, a multitude of problems have plagued these particular analyses such as misestimated standard errors, aggregation bias, and heterogeneity of regression (Burnstein, 1980; Raudenbush & Bryk, 2002). Neither regression analysis was appropriate given the nested structure of data, and hierarchical linear modeling procedures were developed to address these needs.

In the OLS regression with the student as the unit of analysis, students attending the same institution have the same value on each institutional variable. The sharing of the same group can cause dependency among observations. These dependencies may occur because of the shared experiences students have at an institution or because of the way students were initially drawn to an institution. Acknowledging the interdependency of individuals attending the same institution is important because it causes correlations among disturbances, which violates the OLS regression assumption that disturbances are unrelated. In addition, OLS regression cannot capture the positive intra-class correlations that result from the interdependencies among students within the same major or within the same institution and can lead to misestimated standard errors and risk inflation of type I error rates. Furthermore, using institutional variables to predict a student level outcome, such as students' perceived critical thinking ability, forces the researcher to assume that the institution affects all individuals within an institution identically. This is an obvious conceptual error given the ways institutions allocate financial resources.

In the OLS regression with the institution as the unit of analysis, all data were aggregated to the institution level. This introduces the problem of aggregation bias where aggregate relationships generally are much stronger and can be quite different from the relationships in the student level analysis. Moreover, student level variables can change their meaning when aggregated. For instance, in the student level OLS regression model, the scale representing student-faculty interaction measures the time and effort an individual student invests in relationships and interactions with faculty members; whereas, in the institution level OLS regression model, the student-faculty interaction variable represents a characteristic of the normative student body and is the average student-faculty interaction for the institution. While the changes in meaning across levels are not as dramatic in this instance, it could be one reason for the different effects seen in the two regression models. Hence, aggregation bias can have a substantial impact on the substantive interpretations and conclusions drawn from a study.

Hierarchical linear modeling solves the problems associated with the traditional approaches applied in examining college effects on students. First, by acknowledging the multilevel nature of the data, selecting an appropriate unit of analysis is not problematic. Second, hierarchical linear modeling incorporates a unique random effect for each organizational unit in the statistical model and the variability in these random effects is taken into account when estimating standard errors. In other words, the standard errors are adjusted for the intra-class correlation that occurs as a result of the nested data (Raudenbush & Bryk, 2002). Third, hierarchical linear models enable a researcher to estimate a separate set of regression coefficients for each higher level organizational unit and then model variation among the higher level units in their sets of coefficients as multivariate outcomes to be explained by higher level factors, thereby, solving the problem of heterogeneity of regression (Raudenbush & Bryk, 2002). In effect, HLM more accurately reflects the type and structure of data commonly used when studying the influence of college on students.

HLM Estimates of Variance Components

The three-level HLM analysis allows us to partition the total variability in students' perceived critical thinking ability into its three components: among students, among majors within institutions, and among institutions. Calculations of the intra-class correlation coefficients (Table 3) show that 92.4% of the total variance in students' perceived critical thinking is among students, 5.1% is due to differences among majors within institutions, and 2.5% is due to differences among institutions. As can be seen, most of the variability is due to individual differences. Researchers (Burstein, 1980; Ethington, 1997; Kreft & de Leeuw, 1998; Pascarella & Terrenzini, 1991, 2005) studying multilevel structures have observed similar results that most of the variability in hierarchical structures are due to within group differences. An important feature to note concerning the partition of variability is the variation due to differences between majors and differences between institutions. In effect, this shows that majors are more important in explaining variance than institutions and provides evidence of the importance of academic disciplines. Since the HLM analysis allows us to partition the total variability in students' perceived critical thinking ability into its three parts, which is something OLS regression is not able to do, we see a better picture of how the variation in students' perceived critical thinking ability is distributed with the HLM model than with either regression analysis.

For the three-level random-intercept only model used in this study, the variance components to be considered are the proportion reduction in Level-1 residual variance

(σ^2), the proportion reduction in random variation over majors at Level 2 (τ_π), and the proportion reduction in random variation over institutions at Level 3 (τ_β). In this study, the proportion of variance explained in the Level-1 residual variance by the addition of the Level-1 predictors (higher-order thinking skills, academic effort, and student-faculty interaction) is 27.73%. Thus, the student characteristics are explaining 27.73% of the 92.4% of total variation among students. This statistic is calculated by subtracting the Level-1 residual variance from the full Level-1 model (Table 4, σ^2 = 66.193) from Level-1 residual variance from the unconditional model (Table 3, σ^2 = 91.586) then dividing by the Level-1 residual variance from the unconditional model.

The proportion of variance explained in the average critical thinking across majors within institutions by the addition of the Level-2 predictors (hard vs. soft dimension and pure vs. applied dimension) is 21.88%. Thus, the major characteristics are explaining 21.88% of the 5.1% of variability that is due to differences between majors. This statistic is calculated by subtracting the Level-2 residual variance from the full Level-2 model (Table 5, τ_π = 7.422) from the Level-2 residual variance from the full Level-1 model (Table 4, τ_π = 9.500) then dividing by the Level-2 residual variance from the full Level-1 model. An important feature to note is that the variance explained in the average critical thinking across majors within institutions is conditional on the specific Level-1 model, and the variance reduction statistic is only interpretable for models with the same Level-1 model (Raudenbush & Bryk, 2002).

The proportion of variance explained in the average critical thinking across institutions by the addition of the Level-3 predictors (supportive campus environment, some graduate coexistence, high graduate coexistence, primarily commuter, primarily residential, and Barron's ratings of institutional selectivity) is 57.67%. Thus, the institutional characteristics are explaining 57.67% of the 2.5% of the variability that is due to differences between institutions. This statistic is calculated by subtracting the Level-3 residual variance from the full Level-3 model (Table 6, $\tau_\beta = 1.103$) from the Level-3 residual variance from the full Level-2 model (Table 5, $\tau_\beta = 2.606$) then dividing by the Level-3 residual variance from the full Level-2 model. Again, the variance explained in the average critical thinking across institutions is conditional on the specific Level-1 and Level-2 model, and the variance reduction statistic is only interpretable for models with the same Level-1 and Level-2 model.

Estimates of variance explained are not directly comparable between HLM and OLS regression because in the HLM analysis we have taken the total variability in students' perceived critical thinking ability and separated it into its three parts: among students, among majors within institutions, and among institutions. In the HLM analysis, major characteristics can only account for variation among major means. That is, only the parameter variation, τ_π, is explainable. Likewise, institutional characteristics can only account for variation among institutional means. Again, only the parameter variation, τ_β, is explainable. In comparison, ordinary least squares regression employs the total outcome variability to compute the variance explained statistic, R-squared. The variation among students however, reflects individual effects and errors of measurement in the outcome both of which are unexplainable by major characteristics and institutional characteristics in HLM (Raudenbush & Bryk, 2002). Since these statistics are computed in different ways, there is no straightforward comparison of variance explained statistics between OLS regression and HLM analysis. Although variance explained statistics are not directly comparable between analyses, with the HLM analysis we are better able to see how variance is distributed and how variables measured at different levels affects critical thinking.

Comparison of the HLM Fixed Effects to the Two Regression Models

Table 8 gives a comparison of the results across all three analyses. From the results of the HLM estimates of the fixed effects, we see that all three student-level measures have a significant impact on students' perceived critical thinking ability. As in the two regression models, students whose coursework emphasizes higher-order thinking skills perceive greater critical thinking abilities (π_{1jk} = 0.387). Again, like the two regression models, academic effort (π_{2jk} = 0.136) is shown to have a significant, positive influence on students' perceived critical thinking ability. As in the student OLS regression model, student-faculty interaction (π_{3jk} = 0.136) is shown to have a unique influence on students' perceived critical thinking ability. This result is different than what is found in the institution OLS regression model which did not show student-faculty interaction to have significant effect on critical thinking ability. Consistent with the two regression models, results from the HLM analysis show that both major characteristics have a significant effect on the average perceived critical thinking ability across majors. Once more, students majoring in hard disciplines (β_{01k} = 2.655) and students majoring in applied disciplines (β_{02k} = -2.111) perceive greater critical thinking abilities.

In the HLM analysis, the institutional characteristics that have a significant influence on the average critical thinking across institutions are perceptions of supportive campus environment (γ_{001} = 0.140), primarily commuter institutions (γ_{004} = 1.310),

Table 8

Comparison of Results across Analyses

Independent Variables	*Student OLS b (S.E.)*	*Institution OLS b (S.E.)*	*HLM Coefficient (S.E.)*
Higher Order Thinking Skills (HOT)	0.394** (.004)	0.485** (.047)	0.397** (.005)
Academic Effort (AE)	0.144** (.004)	0.192** (.040)	0.136** (.005)
Student-Faculty Interaction (SFI)	0.110** (.004)	0.026 (.041)	0.125** (.005)
HARD	3.028** (.082)	4.047** (.412)	2.655** (.121)
PURE	-2.216** (.072)	-2.726** (.465)	-2.111** (.106)
Supportive Campus Environment (SCE)	0.075** (.005)	0.075** (.009)	0.140** (.010)
Some graduate coexistence (SG)	-0.024 (.102)	-0.057 (.117)	-0.470 (.208)
High graduate coexistence (HG)	0.759** (.129)	0.303 (.252)	-0.054 (.284)
Primarily commuter (PC)	1.291** (.119)	1.328** (.217)	1.31** (.249)
Primarily residential (PR)	0.760** (.096)	0.674** (.170)	0.584* (.204)
Barron's ratings of institutional selectivity (BAR)	0.069 (.037)	0.027 (.073)	0.401** (.080)

$^{*}p < .01$. $^{**}p < .001$.

primarily residential institutions (γ_{005} = 0.584) and selectivity (γ_{006} = 0.401). Findings from the HLM analysis that are analogous to the findings from the two regression models are the conclusions drawn regarding the effects of students' perceptions of supportive campus environment, the residential character of an institution, and some graduate coexistence. As in the two regression models, institutions where students perceive the campus as a supportive and friendly place also tend to report higher average critical thinking scores. In addition, institutions that are primarily commuter or primarily residential tend to have higher average critical thinking scores than institutions that are highly residential. Finally, in all three analyses, the effect of some graduate coexistence is not shown to have a significant relationship with critical thinking ability.

A couple of the effects of the institutional characteristics differ across analyses. Most notably, the results of the HLM analysis demonstrate that institutions that are more selective, as measured by the Barron's rating of institutional selectivity, tend to have higher average critical thinking scores than institutions that are less selective. This is an interesting finding that we do not observe in either regression analysis. Another result that differs across analyses is the effect of high graduate coexistence. In the student OLS regression model, institutions with a high graduate coexistence are shown to have higher critical thinking scores; however, the effect of high graduate coexistence is not significant in either the institution OLS regression model or in the HLM analysis.

Not only did the effects of the independent variables differ across analyses but coefficient estimates differ as well. When comparing coefficient estimates of HLM and OLS regression procedures, Raudenbush and Bryk (2002) make the case that coefficient estimates in HLM will be similar to the estimates in OLS regression, but the estimates of standard errors will tend to be biased downward. They contend that generally the coefficient estimates in the student OLS regression will be more similar to HLM estimates than estimates in the institution OLS regression model, but the degree of agreement between analyses will depend upon the degree of imbalance in the group sample sizes. For instance, if the sample sizes are similar for each higher-level organization, the coefficient estimates will be the similar. If the sample sizes are not similar for all higher-level groups, as it is in this study, coefficient estimates may differ substantially across analyses.

In this study, there was great imbalance in the group sample sizes, which is a common trait in multi-institutional studies. Therefore, we would expect coefficient estimates to vary across analyses, and for the most part they did. One instance where they do not vary as widely is the in the estimates of the student characteristics. Coefficient estimates for the student characteristics were fairly similar across all three analyses. In all three analyses the coefficient estimates for higher order thinking skills and academic effort are essentially the same. The only student characteristic to differ across analyses is the coefficient estimate for student-faculty interaction in the institution OLS regression model.

For the major characteristics, coefficient estimates were consistent across analyses for the pure vs. applied dimension, but varied greatly in the estimate for the hard vs. soft dimension. The estimate for the hard vs. soft dimension in the institution OLS regression model (b = 4.047) is more than one and a half times as large as the HLM estimate (β_{01k} = 2.655). On the other hand, the HLM estimate and the student OLS regression estimate (b = 3.028) are more or less similar across analyses. Again, the coefficient estimates in the

HLM analysis are more similar to the results of the student OLS regression model than the institution OLS regression model.

Across all three analyses, the effects of the institutional characteristics varied widely. The two regression analyses produced identical results for the coefficient estimate for the supportive campus environment scale (b = 0.075); however, the estimate from the HLM analysis (γ_{001} = 0.140) is almost double. While the coefficient estimates of some graduate coexistence are similar across all analyses, essentially no different than zero, the coefficient estimates for high graduate coexistence vary greatly from one analysis to the other. Similarly, the estimates of primarily commuter institutions are similar across analyses, while the estimates of primarily residential institutions vary from one analysis to the other. Finally, the coefficient estimate for selectivity in the HLM analysis (γ_{006} = 0.401) varied greatly from the estimates in the regression models, which are virtually zero.

The differences shown in the major and institutional characteristics can be attributed to the unbalanced nature of the data used in this study. One way to avoid these differences, according to Raudenbush and Bryk (2002), would be to have a similar number of individuals in each group. However, unless a researcher specifically samples equal numbers of individuals in each group, it is rarely the case to find a dataset with an equal number of individuals in each higher-level unit, whether it is an equal number of students in various majors or an equal number of students in multiple institutions. Thus, researchers will typically find that coefficient estimates produced by HLM will differ from the coefficient estimates produced by OLS regression.

As noted above, Raudenbush and Bryk (2002) indicate that estimates of standard errors of the fixed effects will differ across analyses. They maintain that the standard errors produced by the student OLS regression model will generally be too small because OLS regression does not take into account the fact that lower-level units are not independent and are clustered within higher-level units. Nevertheless, this was not the case for the student characteristics. In this study, both the student OLS regression model and the HLM model produced basically the same estimates for the standard errors of the fixed effects for the student characteristics. Given the large sample size (n = 57,276) used in the student OLS regression and the Level-1 HLM model, we would expect the standard errors to be very small, as they were in both analyses. On the other hand, the estimates for the standard errors for the major characteristics and the institutional characteristics in the student OLS regression model are substantially smaller than the HLM estimates, which Raudenbush and Bryk argue will occur. For the Level-2 and Level-3 HLM model, the sample size issue is not as critical because the sample size at these levels are drastically smaller, n = 9,441 and n = 407, respectively. Thus, the results produced by HLM for the standard errors are similar to what Raudenbush and Bryk argue will occur.

Estimates of standard errors in the institution OLS regression vary considerably when compared to the HLM analysis. Standard errors in the institution OLS regression model are higher than the HLM estimates for the student and major characteristics. One reason the standard errors in the institution OLS regression analysis are larger for the student and major characteristics could be due to aggregation bias since aggregate data have stronger correlations and relationships. Another reason for the discrepancy could be due to the differing sample sizes used in the HLM analysis across levels. For the student attributes, the Level-1 HLM analysis uses the sample size at the student-level (*n* = 57,276); thus, since standard errors are a function of sample size, we would expect the standard errors for the student characteristics in the HLM analysis to be substantially smaller than the ones found in the institution OLS regression analysis which used a sample size of *n* = 407. The same is true for the major characteristics. The HLM analysis used a sample size of *n* = 9,441 while the sample size in the institution OLS regression analysis stayed constant (*n* = 407). When we examine the standard errors for the institutional effects, we find what Raudenbush and Bryk (2002) argue will occur: standard errors in the institution OLS regression model are consistently smaller than the standard errors produced by HLM. In this instance, both the HLM analysis and the institution OLS regression analysis are using the same sample size (*n* = 407) to estimate these standard errors. Finally, the estimates of standard errors for the OLS regression analysis with the institution as the unit of analysis are consistently larger than the standard error estimates for the OLS regression analysis with the student as the unit of analysis. Again, one reason the standard errors in the institution OLS regression are larger could be to aggregation bias.

Conclusions

How college affects students is an important topic of research in the higher education literature. Traditionally, higher education researchers have utilized the traditional linear model, ordinary least squares (OLS) regression, to aid in their investigation of the influence of college on students. However, this traditional approach ignores the multilevel nature of the majority of data used in such studies, which can cause a multitude of problems such as misestimated coefficients and standard errors, spurious significant effects, aggregation bias, and heterogeneity of regression. Therefore, a statistical technique that can take into account the multilevel nature of the organization of postsecondary education, such as hierarchical linear modeling (HLM), is needed.

In this study, I examined the influences on seniors' perceived critical thinking ability three ways in order to illustrate the differing conclusions one may come to depending upon the type of analysis chosen. The first approach was an OLS regression with the student as the unit of analysis, which is generally the statistical approach taken by a majority of higher education scholars. The second approach was an OLS regression with the institution as the unit of analysis, which is generally seen as an alternative to the student OLS regression model. The third approach was a three level hierarchical linear model with student attributes modeled at Level 1, characteristics of the academic disciplines modeled at Level 2, and characteristics of the institutions modeled at Level 3. Thus, a statistical approach that takes into account the multilevel nature of the organization of postsecondary education. Overall, evidence from this study demonstrates that one would come to substantively different conclusion regarding the influences on students' perceived critical thinking ability depending on the type of analysis chosen, especially in regards to the effects of the institutional characteristics.

The findings of this study can be summed up as follows. First, the results of the institution OLS regression model cannot be considered reliable. Findings from the institution OLS regression model differed substantially from the results of the other two analyses. In the institution OLS regression model, student-faculty interaction and selectivity were not found to have a significant relationship with the average students' perceived critical thinking ability as was found in the HLM analysis. These are two important findings that are not illustrated in the institution OLS regression analysis. If a researcher would have performed this analysis, he or she would have concluded that institutions that foster environments that lead to greater interactions among faculty members and students does not have a significant impact the average students' perceived critical thinking ability. This is contrary to the abundant literature (Astin, 1993; Kuh & Hu, 2001; Pace, 1979, 1984; Tinto, 1987) that has demonstrated the importance of student-faculty interaction on students' growth and development in college. If a researcher had performed this analysis, he or she would have also concluded that the selectivity of an institution is not related to the average students' perceived critical thinking ability. However, results from the HLM analysis tend to suggest otherwise. In addition to the different substantive conclusions, the coefficient estimates and standard errors in the institution OLS regression analysis differed substantially from the coefficient estimates and standard errors in the HLM analysis. With such contradictory findings in the institution OLS regression analysis, it is expected that one would not come to accurate conclusions regarding the influences on the average students' perceived critical thinking ability with an OLS regression with the institution as the unit of analysis.

Second, the findings from the student OLS regression model can only be partially trusted. Evidence from this study suggests that one can be fairly confident in the results obtained for the student and major characteristics. Even when modeling major and institutional characteristics in the regression model, one can still trust the results of the student-level variables and the major-level variables. In addition, the coefficient estimates for the student characteristics and major characteristics are similar to those found in the

HLM analysis. Estimates of standard errors are similar for the student characteristics but differ for the major characteristics. Thus, if a researcher had performed an OLS regression with the student as the unit of analysis, the researcher would have come to the same conclusions regarding the effects of the student characteristics as he or she would have if an HLM analysis were performed but risk inflation of type I error rates for the major characteristics.

Where the student OLS regression analysis and the HLM analysis primarily differ are in the effects of the institutional characteristics. In the student OLS regression model, I argue that the effects of the institutional characteristics are minimal at best given the large sample size and relatively small standardized coefficients. Thus, if a researcher was to perform the OLS regression with the student as the unit of analysis, he or she would erroneously conclude that the institutional characteristics do not have a significant impact on students' perceived critical thinking ability; thus, concluding that different types of four-year institutions have essentially the same impact on students' perceived critical thinking ability. Furthermore, coefficient estimates and standard errors for the institutional characteristics in the student OLS regression model were substantially smaller than those in the HLM analysis as expected according the Raudenbush and Bryk (2002).

Third, when institutional effects are of prime importance, one should perform an HLM analysis in order to be confident in the results obtained for the institutional effects. As discussed earlier, the results from both OLS regression analyses failed to accurately describe the effects of the institutional characteristics. Thus, when a researcher is interested in institutional effects, which is often the case when studying college effects on students, researchers need to utilize HLM procedures in order to be confident in the results. Ordinary least squares regression has been the foundation on which college effects studies have been built. However, evidence from this sample suggests that ordinary least squares regression is not capable of accurately detecting institutional effects in the presence of multilevel data. Given the discrepancy in results across all three analyses and the lack of consistency in the literature involving the influence of college on students (Pascarella & Terenzini, 1991, 2005), regular use of hierarchical linear modeling may be one way to yield more valid and informative findings in the college effects literature.

The primary interest of this study was to investigate the differences in substantive conclusions one may come to depending upon the type of analysis chosen, OLS regression or HLM. Thus, empirical data were employed in this study. By using empirical data, we are dealing with a more realistic research situation instead of a robustness study where data are computed based on fixed parameters then altered to meet certain criteria. Using empirical data, instead of data computed based on certain parameters, places this study in the literature of college impact studies, and in doing so, we are better able to test the theoretical framework from the higher education literature. When we use OLS regression, the statistical model does not fit the nature of the data used when investigating the influence of college on students. On the other hand with HLM, the statistical model fits the theoretical model where students are nested within majors nested within institutions. Thus, by using empirical data, we were able to examine whether a misspecified statistical model, such as OLS regression, can produce parameter estimates comparable to a statistical model that better represents the theoretical model under study, such as HLM.

Since this is empirical data, we are unable to know the true parameter estimate. However, if we acknowledge that HLM provides the best statistical model, and as a result, gives us the best parameter estimates, we can investigate how parameter estimates produced by HLM compare with the estimates produced by OLS regression. So how do parameter estimates compare across analyses? Evidence from this sample suggests that OLS regression is limited in its ability to produce accurate parameter estimates. As discussed earlier, the OLS regression with the institution as the unit of analysis produced parameter estimates that were substantially different than those produced by the HLM analysis. Therefore, researchers should use caution when using an OLS regression with the institution as the unit of analysis to study college effects on students. The OLS regression with the student as the unit of analysis produced parameter estimates similar to those found in the HLM analysis for the student and major characteristics; however, the parameter estimates for the institutional characteristics differed considerably. Therefore, when using an OLS regression with the student as the unit of analysis to study the influence of college on students, researchers can be fairly confident in the parameter estimates of the lower-level variables, such as estimates for student characteristics, but must be cautious when interpreting the parameter estimates for the higher-level variables, such as institutional characteristics.

In regards to the findings of this study, I make the following recommendations concerning the appropriate analysis in the presence of multilevel data. First, if a researcher has only collected data on students, yet still recognizes the multilevel nature of the data, one would come to similar conclusions with HLM and OLS regression. Second, if a researcher has data collected on multiple levels, i.e., student characteristics, major characteristics, and institutional characteristics, results from OLS regression and HLM will differ in regards to higher-order variables. The researcher can be fairly confident in their findings regarding the lower-level variables but cannot trust findings regarding higher-level variable. In this study, the student OLS regression and the HLM analysis produced similar results for the student attributes and major characteristics but produced substantively different results for the institutional effects. With this in mind, I would caution researchers in their attempt to use ordinary least squares regression to discern relationships between institutional variables. Given that hierarchical linear modeling more accurately describes the nature of data under investigation, when data are collected at multiple levels, and when sample size is adequately large enough, hierarchical linear modeling yields the best parameter estimates and can allow for a richer, more thorough investigation of the phenomenon under study.

REFERENCES

Astin, A. W. (1993). *What matters in college? Four critical years revisited*. San Francisco: Jossey-Bass Publishers.

Biglan, A. (1973a). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology, 57*(3), 195-203.

Biglan, A. (1973b). Relationships between subject matter characteristics and the structure and output of university departments. *Journal of Applied Psychology, 57*(3), 204-213.

Bohrnstedt, G. W., & Carter, T. M. (1971). Robustness in regression analysis. *Sociological Methodology, 3,* 118-146.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. In D. C. Berliner (Ed.), *Review of Research in Education* (Vol. 8). Washington, DC: American Educational Research Association.

Burstein, L., Linn, R. L., & Capell, F. J. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journals of Educational Statistics, 3,* 347-383.

Burstein, L., & Miller, M. D. (1980). Regression-based analyses of multi-level educational data. *New Directions for Methodology of Social and Behavioral Sciences, 6*, 194-211.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

de Leeuw, J., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics, 11*(1), 57-85.

de Leeuw, J., & Kreft, I. G. G. (1995). Questioning multilevel models. *Journal of Educational and Behavioral Statistics, 20(*2), 171-189.

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

Draper, D. (1995). Inference and hierarchical modeling in the social sciences. *Journal of Educational and Behavioral Statistics, 20*(2), 115-147.

Ethington, C. A. (1997). A hierarchical linear modeling approach to studying college effects. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research* (Vol. 12, pp. 165-194). New York: Agathon Press.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland, 15,* 246–263.

Gauss, C. F. (1809). *Theory of the motion of the heavenly bodies moving about the sun in conic sections.* (C. H. Davis, Trans.) Boston: Little, Brown and Company (1857).

Hannan, M., & Burstein, L. (1974). Estimation from grouped observations. *American Sociological Review, 39*, 374-392.

Hox, J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp. 147-154). New York: Springer Verlag.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Kim, K. S. (1990). *Multilevel data analysis: A comparison of analytical alternatives.* Unpublished doctoral dissertation, University of California, Los Angeles.

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling.* Newbury Park, CA: Sage.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. (1995). The effects of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30*(1), 1-22.

Kuh, G. D. (2001). *The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties*. Bloomington: IN: Indiana University Center for Postsecondary Research.

Kuh, G. D., & Hu, S. (2001). The effects of student-faculty interaction in the 1990s. *The Review of Higher Education, 24*(3), 309-332.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B, 34*, 1-41.

Maas, C. J., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58*(2), 127-137.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86-92.

Pace, C. R. (1979). *Measuring outcomes of college: Fifty years of findings and recommendations for the future*. San Francisco: Jossey-Bass.

Pace, C. R. (1984). *Measuring the quality of college student experiences. An account of the development and use of the College Student Experiences Questionnaire*. Los Angeles: Higher Education Research Institute.

Pascarella, E., & Terenzini, P. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco: Jossey-Bass.

Pascarella, E., & Terenzini, P. (2005). *How college affects students: A third decade of research*. San Francisco: Jossey-Bass.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research, third edition*. New York: Holt, Rinehart and Winston.

Pike, G., Kuh, G., & McCormick, A. (2008, November). Learning community participation and educational outcomes: Direct, indirect and contingent relationships. Paper presented at the 2008 Association for the Study of Higher Education Conference in Jacksonville, FL.

Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education, 59*(1), 1-17.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods, second edition*. Thousand Oaks, CA: Sage Publications.

Raudenbush, S.W., Bryk, A.S., & Congdon, R. (2000). HLM 5 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *Sociological Review, 15*(3), 351-357.

Smart, J. C. (2005). Perspectives of the editor: Attributes of exemplary research manuscripts employing quantitative analyses. *Research in Higher Education, 46*(4), 461-477.

Smith, A. F. M. (1973). A general Bayesian linear model. *Journal of the Royal Statistical Society, Series B,* 35, 61-75.

Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition.* Chicago: The University of Chicago Press.

Toutkoushian, R. K., & Smart, J. C. (2001). Do institutional characteristics affect student's gains from college? *The Review of Higher Education, 25*(1), 39-61.

Webster, W. J., Mendro, R. L., Orsak, T. H., & Weerasinghe, D. (April, 1996). The applicability of selected regression and hierarchical linear models to the estimation of school and teacher effects. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York.

Appendix A

Items comprising the variables used in the analyses and the construction of scales

---

Critical Thinking (CT) (α = 0.79)

Computed by summing across the following three items then converting to a T score:

- To what extent has your experience at this institution contributed to your knowledge, skills and personal development in thinking critically and analytically?
- To what extent has your experience at this institution contributed to your knowledge, skills and personal development in analyzing quantitative problems?
- To what extent has your experience at this institution contributed to your knowledge, skills and personal development in solving complex real-world problems?

Each item is coded 1 = very little, 2 = some, 3 = quite a bit, 4 = very much.

STUDENT MEASURES

Course Emphasis on Higher-Order Thinking Skills (HOT) (α = 0.80)

Computed by summing across the following six items then converting to a T score:

- How much as your coursework emphasized analyzing the basic elements of an idea, experience, or theory, such as examining a particular case or situation in depth and considering its components?
- How much as your coursework emphasized synthesizing and organizing ideas, information, or experiences into new, more complex interpretations and relationships?
- How much as your coursework emphasized making judgments about the value of information, arguments, or methods, such as examining how others gathered and interpreted data and assessing the soundness of their conclusion?
- How much as your coursework emphasized applying theories or concepts to practical problems or in new situations?
- How often have you worked on a paper or project that required integrating ideas or information from various sources?
- How often have you put together ideas or concepts from different courses when completing assignments or during class discussions?

Each item is coded same as above.

Academic Effort (AE) (α = 0.67)

Computed by summing across the standardized scores of the following eleven items then converting to a T score:

- How often have you prepared two or more drafts of a paper or assignment before turning it in?
- How often have you worked harder than you thought you could to meet an instructor's standards or expectations?

Each of these two items are coded 1 = never, 2 = sometimes, 3 = often, 4 = very often.

- During the current school year, about how much reading and writing have you done?
    - Number of assigned textbooks, books, or book-length packs of course readings.
    - Number of written papers or reports of 20 pages or more.
    - Number of written papers or reports of between 5 and 19 pages.
    - Number of written papers or reports of fewer than 5 pages.

Each of these four items are coded 1 = None, 2 = 1-4, 3 = 5-10, 4 = 11-20, 5 = More than 20.

- Mark the extent to which your examinations during the current school year have challenged you to do your best work.

This item is coded 1 = very little to 7 = very much.

- In a typical week, how many homework problem sets do you complete?
    - Number of problem sets that take you more than an hour to complete.
    - Number of problem sets that take you less than an hour to complete.

Each of these two items are coded 1 = None, 2 = 1-2, 3 = 3-4, 4 = 5-6, 5 = More than 6.

- How many hours a week do you spend preparing for class (studying, reading, writing, doing homework or lab work, analyzing data, rehearsing, and other academic activities)?

This item is coded 1 = 0 hours, 2 = 1-5 hours, 3 = 6-10 hours, 4 = 11-15 hours, 5 = 16-20 hours, 6 = 21-25 hours, 7 = 26-30 hours, 8 = more than 30 hours.

- To what extent does your institution emphasize spending significant amounts of time studying and on academic work?

This item is coded 1= very little, 2 = some, 3 = quite a bit, 4 = very much.

Student-Faculty Interaction (SFI) ($\alpha = 0.77$)

Computed by summing across the following five items then converting to a T score:

- How often have you used e-mail to communicate with an instructor?
- How often have you discussed grades or assignments with an instructor?

- How often have you talked about career plans with a faculty member or advisor?
- How often have you discussed ideas from your readings or classes with faculty members outside of class?
- How often have you worked with a faculty member on activities other than coursework (committees, orientation, student life activities, etc.)?

Each is coded 1 = never, 2 = sometimes, 3 = often, 4 = very often.

MAJOR CHARACTERISTICS

Hard vs. Soft (HARD)

A dichotomous variable coded 0 = soft, 1 = hard.

Pure vs. Applied (PURE)

A dichotomous variable coded 0 = applied, 1 = pure.

INSTITUTIONAL CHARACTERISTICS

Supportive Campus Environment (SCE) ($\alpha = 0.78$)

Computed by summing across the standardized scores of the following six items then converting to a T scores:

- Quality of your relationships with other students.

Item ranges from 1 = unfriendly, unsupportive, sense of alienation to 7 = friendly, supportive, sense of belonging.

- Quality of your relationships with faculty members.

Item ranges from 1 = unavailable, unhelpful, unsympathetic to 7 = available, helpful, sympathetic.

- Quality of your relationships with administrative personnel and offices.

Item ranges from 1 = unhelpful, inconsiderate, rigid to 7 = helpful, considerable, flexible.

- To what extent does your institution emphasize providing the support you need to help you succeed academically?
- To what extent does your institution emphasize helping you cope with your non-academic responsibilities (work, family, etc.)?
- To what extent does your institution emphasize providing the support you need to thrive socially?

Each is coded 1= very little, 2 = some, 3 = quite a bit, 4 = very much. Since this is considered an institutional characteristic, it was aggregated for each institution.

Graduate Coexistence

Values in this variable were merged into three categories: no graduate coexistence, some graduate coexistence, and high graduate coexistence. Then, two dummy variables were created. One (SG) was coded 1 = some graduate coexistence, 0 = otherwise. The other (HG) was coded 1= high graduate coexistence, 0 = otherwise.

Residential Character

Values in this variable will be merged into three categories: primarily commuter, primarily residential, and highly residential. Then, two dummy variables will be created. One (PC) was coded 1 = primarily commuter, 0 = otherwise. The other (PR) was coded 1 = primarily residential, 0 = otherwise.

Barron's Ratings of Institutional Selectivity (BAR)

Has eleven categories ranging from 1 = noncompetitive to 6 = most competitive.

Appendix B

List of Majors and Biglan (1973a, 1973b) classification

Major	*Pure vs. Applied*	*Hard vs. Soft*
Art, fine and applied	Pure	Soft
English (language and literature)	Pure	Soft
History	Pure	Soft
Journalism	Applied	Soft
Language and literature (except English)	Pure	Soft
Music	Pure	Soft
Philosophy	Pure	Soft
Speech	Applied	Hard
Theater or drama	Pure	Soft
Theology or religion	Applied	Soft
Biology (general)	Pure	Hard
Biochemistry or biophysics	Pure	Hard
Environmental science	Pure	Hard
Microbiology or bacteriology	Pure	Hard
Zoology	Pure	Hard
Accounting	Applied	Soft
Business administration (general)	Applied	Soft
Finance	Applied	Soft
Marketing	Applied	Soft
Management	Applied	Soft
Business education	Applied	Soft
Elementary/middle school education	Applied	Soft
Music or art education	Applied	Soft
Physical education or recreation	Applied	Soft
Aero-/astronautical engineering	Applied	Hard
Civil engineering	Applied	Hard
Chemical engineering	Applied	Hard
Electrical or electronic engineering	Applied	Hard
Industrial engineering	Applied	Hard
Materials engineering	Applied	Hard
Mechanical engineering	Applied	Hard
General/other engineering	Applied	Hard
Atmospheric science (including meteorology)	Pure	Hard
Chemistry	Pure	Hard
Earth science (including geology)	Pure	Hard
Mathematics	Pure	Hard
Physics	Pure	Hard
Statistics	Pure	Hard
Architecture	Applied	Soft
Urban planning	Applied	Soft
Medicine	Applied	Hard
Nursing	Applied	Soft

Major	*Pure vs. Applied*	*Hard vs. Soft*
Pharmacy	Applied	Hard
Allied health/other medical	Applied	Soft
Anthropology	Pure	Soft
Economics	Applied	Soft
Ethnic studies	Pure	Soft
Geography	Pure	Soft
Political science	Pure	Soft
Psychology	Pure	Soft
Social work	Applied	Soft
Sociology	Pure	Soft
Agriculture	Applied	Hard
Communications	Applied	Soft
Family Studies	Applied	Soft
Kinesiology	Pure	Hard
Criminal justice	Applied	Soft
Public administration	Applied	Soft

Appendix C

Correlations, Means, and Standard Deviation for Student-level OLS Regression

	CT	HOT	AE	SFI	HARD	PURE	SCE	SG	HG	PC	PR	BAR
CT	1.000											
HOT	0.509	1.000										
AE	0.368	0.463	1.000									
SFI	0.355	0.503	0.390	1.000								
HARD	0.100	-0.035	0.009	-0.017	1.000							
PURE	-0.075	0.027	0.017	0.072	0.148	1.000						
SCE	0.076	0.089	0.103	0.168	-0.062	0.103	1.000					
SG	-0.027	-0.020	-0.016	0.007	-0.085	-0.067	0.101	1.000				
HG	0.015	-0.037	-0.066	-0.103	0.111	-0.094	-0.418	-0.642	1.000			
PC	-0.015	-0.049	-0.068	-0.115	-0.021	-0.071	-0.382	-0.045	0.265	1.000		
PR	-0.003	-0.043	-0.042	-0.034	0.013	-0.113	-0.081	0.185	0.064	-0.403	1.000	
BAR	0.038	0.082	0.068	0.061	0.069	0.147	0.119	-0.269	0.134	-0.358	-0.113	1.000
Means	50.051	50.182	50.111	50.090	0.246	0.488	47.599	0.535	0.264	0.241	0.338	3.683
St. dev.	9.943	9.924	9.914	9.986	0.431	0.500	8.582	0.499	0.441	0.428	0.473	1.121

Appendix D

Correlations, Means, and Standard Deviation for Institution-level OLS Regression

	CT	HOT	AE	SFI	HARD	PURE	SCE	SG	HG	PC	PR	BAR
CT	1.000											
HOT	0.601	1.000										
AE	0.523	0.671	1.000									
SFI	0.391	0.670	0.595	1.000								
HARD	0.365	0.023	0.060	-0.083	1.000							
PURE	0.051	0.528	0.414	0.522	-0.177	1.000						
SCE	0.438	0.448	0.461	0.650	-0.093	0.292	1.000					
SG	-0.079	-0.115	-0.122	-0.025	-0.221	-0.249	0.066	1.000				
HG	-0.016	-0.182	-0.257	-0.380	0.290	-0.196	-0.416	-0.547	1.000			
PC	-0.034	-0.227	-0.279	-0.488	-0.035	-0.243	-0.363	0.027	0.247	1.000		
PR	0.008	-0.107	-0.135	-0.034	-0.027	-0.256	-0.014	0.211	0.014	-0.389	1.000	
BAR	0.156	0.377	0.308	0.264	0.180	0.428	0.109	-0.269	0.108	-0.357	-0.130	1.000
Means	50.024	50.270	50.224	50.549	0.232	0.498	48.961	0.593	0.170	0.244	0.319	3.426
St. dev.	2.105	2.169	2.268	2.726	0.168	0.191	9.068	0.492	0.376	0.430	0.466	1.100