2014

Reporting Methods and Analyses in Higher Education Research: Hierarchical Linear and OLS Regression Models

Kimberly Quandt Fath
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Part of the Higher Education Administration Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 2014 Kimberly Quandt Fath
LOYOLA UNIVERSITY CHICAGO

REPORTING METHODS AND ANALYSES IN HIGHER EDUCATION RESEARCH:
HIERARCHICAL LINEAR AND OLS REGRESSION MODELS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

PROGRAM IN HIGHER EDUCATION

BY
KIMBERLY QUANDT FATH
CHICAGO, IL
DECEMBER 2014
ACKNOWLEDGMENTS

While this work is entirely my own, this dissertation would not have been completed without the support of many individuals and groups.

I am immensely grateful to the current and former faculty at Loyola University Chicago who have encouraged, challenged, and educated me. These include Drs. Terry Williams and Jennifer Grant Haworth, who helped me acquire a foundation in scholarship of this field and to Dr. Lisa Lattuca who challenged me to think more deeply about the ways in which knowledge is constructed and the forces that act on it. Her influence is a thread throughout this work.

I want to thank the members of my committee, whose expertise and guidance helped shape the focus of this dissertation: Drs. Mark Engberg, Terri Pigott, and John Dugan. They offered encouragement at the times I needed it most and reminded me that my research is important. Dr. Dugan’s research using data from the Multi-Institutional Study of Leadership served as inspiration for my topic. Dr Pigott’s expert knowledge of research methodology and multilevel analyses offered important insights into the application of the technique.

There are many others who have supported me along this journey. I want to thank the librarians at Duke University Libraries and the School of Library Science at the University of North Carolina at Chapel Hill for introducing me to the scholarship on bibliometrics; my colleagues at Elon University who offered disciplinary perspectives
that informed this work; friends at UNC-Chapel Hill who encouraged me to stick with it when the “going got tough;” the virtual communities at #sadoc and #phdchat who helped me to see the humor in the challenges of the dissertation process; and my fellow students at Loyola University Chicago who encouraged and commiserated from afar. A special thank you goes to Robyn Silver and Debbie Stockert. I still miss our coffee dates and am grateful for your unconditional support of this work.

I want to acknowledge my family who has supported me in more ways than even I realize. My mother, Holly Lee, and siblings, Terry, Kevin, and Heidi Quandt, and Marianne Lowell – thanks for the laughs and keeping me grounded when I started to stray. I have to thank my children, Kira, Emma, and Michael Fath. I hope you never forget that some things are still worth doing – even when they are difficult. I learned that from your great grandfather, Arthur Quandt and hope I’ve shared that lesson with you.

Finally, I have to thank my husband and partner, Michael Fath. You supported me in more ways than I thought possible. I will forever be in your debt.
To my husband and partner, Michael Fath,
and my children,
Kira, Emma, and Michael
# TABLE OF CONTENTS

## ACKNOWLEDGEMENTS

iii

## LIST OF TABLES

ix

## LIST OF FIGURES

xi

## ABSTRACT

xii

## CHAPTER ONE: PROBLEM AND CONTEXT

1

- Background of the Study
  - Persistent Use of OLS Regression
  - National Data Set Studies
- Statement of the Problem
- Purpose of the Study
- Research Questions
- Nature of the Study
- Significance of the Study
- Definition of Terms
  - Multilevel Data
  - Multilevel Analysis
  - Multilevel Model
  - Hierarchical Linear Models
  - HLM/HLM6/HLM7
- Chapter Summary

## CHAPTER TWO: REVIEW OF RELEVANT LITERATURE

19

- Evolution of Multilevel Analysis
  - The Problem of Multilevel Data
  - Statistical Evidence
  - Applying Multilevel Analysis to Education Research
- Building and Reporting Hierarchical Linear Models
  - The Sample
  - Model Development and Specification
  - Data Preparation
  - Estimation Techniques
  - Model Building
  - Reporting Results
- Hierarchical Linear Models as the Focus of Research
  - Higher Education Literature about Multilevel Analyses
  - Higher Education Literature Using Multilevel Data
- Chapter Summary
### CHAPTER THREE: RESEARCH METHODS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Analysis</td>
<td>62</td>
</tr>
<tr>
<td>Units of Analysis and Sampling</td>
<td>63</td>
</tr>
<tr>
<td>Coding Procedures</td>
<td>64</td>
</tr>
<tr>
<td>Analyses and Reporting</td>
<td>66</td>
</tr>
<tr>
<td>Study Design</td>
<td>67</td>
</tr>
<tr>
<td>HLM Studies Citation Analysis</td>
<td>68</td>
</tr>
<tr>
<td>HLM Studies Content Analysis</td>
<td>73</td>
</tr>
<tr>
<td>Regression Studies Content Analysis</td>
<td>78</td>
</tr>
<tr>
<td>Trustworthiness of the Data and Analyses</td>
<td>81</td>
</tr>
<tr>
<td>Ethical Issues Related to this Research</td>
<td>82</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>83</td>
</tr>
</tbody>
</table>

### CHAPTER FOUR: RESULTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Characteristics</td>
<td>84</td>
</tr>
<tr>
<td>Sample One: HLM Studies</td>
<td>85</td>
</tr>
<tr>
<td>Sample Two: Regression Studies</td>
<td>86</td>
</tr>
<tr>
<td>Results</td>
<td>87</td>
</tr>
<tr>
<td>Research Question One: HLM Studies Citation Analysis</td>
<td>89</td>
</tr>
<tr>
<td>Research Question Two: HLM Studies Content Analysis</td>
<td>99</td>
</tr>
<tr>
<td>Research Question Three: Regression Studies Content Analysis</td>
<td>120</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>131</td>
</tr>
</tbody>
</table>

### CHAPTER FIVE: SUMMARY AND DISCUSSION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevant Literature</td>
<td>132</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>133</td>
</tr>
<tr>
<td>Review of Methods</td>
<td>134</td>
</tr>
<tr>
<td>Summary of Results</td>
<td>136</td>
</tr>
<tr>
<td>HLM Studies Sample Citation Analysis</td>
<td>137</td>
</tr>
<tr>
<td>HLM Studies Sample Content Analysis</td>
<td>138</td>
</tr>
<tr>
<td>Regression Studies Sample Content Analysis</td>
<td>139</td>
</tr>
<tr>
<td>Discussion of Results</td>
<td>143</td>
</tr>
<tr>
<td>The Literature of Hierarchical Linear Models: Expanding our Method</td>
<td>146</td>
</tr>
<tr>
<td>The Belief that OLS Regression and Hierarchical Linear Models are</td>
<td></td>
</tr>
<tr>
<td>Equivalent</td>
<td></td>
</tr>
<tr>
<td>The Effect of Journal Guidelines</td>
<td>147</td>
</tr>
<tr>
<td>Implications and Recommendations</td>
<td>151</td>
</tr>
<tr>
<td>Selection of Analytic Technique</td>
<td>154</td>
</tr>
<tr>
<td>Reporting Results</td>
<td>159</td>
</tr>
<tr>
<td>Strengthening the Tools of Our Research</td>
<td>162</td>
</tr>
<tr>
<td>Limitations</td>
<td>165</td>
</tr>
<tr>
<td>Recommendations for Future Research</td>
<td>169</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>172</td>
</tr>
</tbody>
</table>

vii
APPENDIX A: HLM STUDIES SAMPLE SOURCES 175
APPENDIX B: HLM STUDIES CITATION ANALYSIS CODE BOOK 182
APPENDIX C: HLM STUDIES CONTENT ANALYSIS CODE BOOK 186
APPENDIX D: REGRESSION STUDIES SAMPLE SOURCES 191
APPENDIX E: REGRESSION STUDIES CONTENT ANALYSIS CODE BOOK 197
APPENDIX F: SAMPLE SCREENING PROCEDURES 202
APPENDIX G: HLM STUDIES SAMPLE CITATION ANALYSIS REFERENCE LIST 207
APPENDIX H: FREQUENCY COUNT FOR HLM STUDIES SAMPLE SOURCES CITED 217
APPENDIX I: HLM STUDIES SAMPLE DATA SOURCES 221
REFERENCE LIST 223
VITA 237
LIST OF TABLES

Table 1. Contributions to the Statistical Development of Multilevel Analyses ..........25
Table 2. Guiding Questions Used to Create HLM Reporting Codebook .....................76
Table 3. Distribution of HLM Studies Sample Articles by Journal Source ..................86
Table 4. Distribution of Regression Studies Sample Articles by Journal Source ..........88
Table 5. Sources Cited by at Least Three Studies in the HLM Studies Sample ............93
Table 6. Types of Sources Cited in HLM Studies Sample Articles ........................94
Table 7. Journal Source Frequency of Articles Cited in HLM Studies Sample Articles ..........................................................................................................................95
Table 8: Classification of Citation Use in Text ..........................................................97
Table 9. HLM Studies Sample by Type of Multilevel Model ....................................103
Table 10. HLM Studies Sample Sources Reported Greater than Two Times ............104
Table 11. Sampling Strategies Represented in HLM Studies Sample .......................106
Table 12. Justification of Hierarchical Linear Model .................................................108
Table 13. Management of Missing Data in HLM Studies Sample .............................109
Table 14. Justification for Predictor Selection or Retention in HLM Studies Sample ....112
Table 15. Methods of Reporting Fixed and Random Effects in HLM Studies Sample ..117
Table 16. Levels of Predictors in Regression Studies Sample ....................................124
Table 17. Articles Providing Justification for Analytic Technique in Regression Studies Sample ..................................................................................................................................................124
Table 18. Arguments Presented to Justify Single-Level Model in Regression Studies Sample ..................................................................................................................................................125
Table 19. Adjustments to Model Building Process Regression Studies Sample ...........129
Table 20. Distribution of Adjustments to Modeling Process in Regression Studies Sample

Table 21. Results of HLM Studies Sample Construction Screening
LIST OF FIGURES

Figure 1. Distribution of HLM Studies Sample Articles by Publication Year ..................87

Figure 2. Distribution of Regression Studies Sample Articles by Publication Year ........88
ABSTRACT

This study examined the reporting practices used by higher education scholars to communicate the methods and analyses of studies using hierarchical linear and OLS regression models. The study consisted of three parts: (a) a citation analysis of sources associated with methods of hierarchical linear models, (b) a content analysis of reporting practices associated with studies that used hierarchical linear models, and (c) a content analysis of the arguments and corrections used by scholars who have used OLS regression techniques on nested data. When possible, results were compared to similar research. The data for this research was drawn from *Journal of Higher Education, Review of Higher Education, Journal of College Student Development*, and *Research in Higher Education*.

This study provides some of the first empirical evidence of how scholars communicate the tools of research in the published literature of the field. It also identified the reasons scholars use when electing to apply OLS regression techniques on data and designs for which multilevel approaches may also be suitable. Findings suggest a need to examine the current practices of the field and to identify a model of best practices for communicating and reporting the methods and results of studies using hierarchical linear models.
CHAPTER ONE
PROBLEM AND CONTEXT

Higher education research studies frequently use analyses of national survey data to inform practice, policy, and theory (Astin, 1993; Dugan & Komives, 2007; Pike & Kuh, 2005). These surveys often use complex sampling designs which cluster individuals within institutions as well as modeling techniques that use predictors measured at different levels (Dugan & Komives, 2007; Kuh, 2009; Wine, Cominole, Wheeless, Dudley, & Franklin, 2006). The selection of a multilevel data structure may reflect both theoretical and practical considerations (Ethington, 1997). Complex samples can be constructed using sampling techniques that are random at one or both levels (Hox, 1995, 1998). Modeling outcomes using these data sets create additional challenges to scholars. Despite violating a key assumption of the technique, scholars seeking to model outcomes have used ordinary least squares (OLS) regression. This analytic method identifies and measures the level of association between independent variables and the outcome measure (Pedhazur, 1997).

The combination of OLS regression techniques and complex data sets is less than statistically ideal as the sampling design violates a key assumption of OLS regression, that the units of analysis are independent (Ethington, 1997). This increases the probability of Type I error, defined as the rejection of the null hypothesis when it is actually true, due to underestimated standard errors (Thomas & Heck, 2001). In
addition, theories have emerged that incorporate predictors that reflect the multilevel nature of the data. There is an alternative to using OLS regression to model multilevel data. Dating to the late 1970s, statisticians developed analytic methods that can model multilevel data with better statistical accuracy (Gelman & Hill, 2007). Multilevel modeling separates and evaluates differences between groups as well as differences within groups (Ethington, 1997; Gelman & Hill, 2007). The technique was introduced to higher education researchers in the late 1990s and began to appear in peer-reviewed journals in the early 2000s.

This research used the text from published studies to examine a series of related questions to understand how higher education, as a knowledge community, reports and analyzes multilevel data. Using literature from four major journals of the field, this dissertation answers the following questions:

1. What methodological sources have been cited by higher education scholars who have published studies that used hierarchical linear models?

2. Using Dedrick et al.’s (2009) analytic framework for examining the narrative content of studies using hierarchical linear modeling, what methodological issues were included/omitted in narratives of studies using the technique?

3. What reasons do scholars give in published articles to justify the use of single-level modeling approaches on complex data?

There is no comprehensive synthesis of higher education research that reports the results of studies using hierarchical linear models. The results of this study provide this synthesis and offer insight into perceptions and misperceptions about hierarchical linear
models. These findings make it possible to identify guidelines scholars may use when selecting an analytic technique, modeling multilevel data, or building models with predictors measured at different levels.

**Background of the Study**

Beginning in the 1970s, advances in statistical theory and procedure have resulted in the development of techniques to correct the problems of using OLS regression on nested data sets. Thomas and Heck (2001) identified two classes of solutions to the problems associated with modeling multilevel data. The first class of solutions were design-based consisting of corrections to OLS regression procedures while maintaining a single level of analysis. The second class of solutions consisted of modeling the data using multilevel techniques.

Thomas and Heck (2001) noted four types of design-based corrections. These corrections included the use of specialized software or procedures, adjusting estimated standard errors using a design effect, modifying weighting to change the effective sample size, or using a more conservative $p$-value in hypothesis tests. Model-based solutions, known as multilevel modeling, are analyses to account for variation at each level of the data set. Thomas and Heck indicated that using multilevel modeling was the preferred approach when a researcher seeks to account for both group and individual level effects. In addition to producing estimates for each of the independent variables, multilevel analyses partition total variance into within- and between- group components. The technique is also appropriate for a variety of questions in higher education research.
including longitudinal designs (Hausmann, Schofield, & Woods, 2007) and meta-analyses (Denson & Seltzer, 2011).

Applied statisticians established, mathematically, that multilevel analysis provides more accurate results for multilevel data (Burstein, 1980; Gelman & Hill, 2007; Hox & Kreft, 1994). The procedures for producing these estimates are complex and, in the early years of the theoretical approaches not practical to perform manually. Over the past several years, statistical software has been developed to perform various types of multilevel analyses. The computer program used most frequently by higher education researchers for multilevel modeling is HLM: Hierarchical Linear and Nonlinear Modeling (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011a).

Persistent Use of OLS Regression

In the early 2000s, few scholars mentioned the problems associated with OLS regressions of multilevel data. By the mid to late 2000s, authors acknowledged the potential problems with nested data and OLS regression, but argued that there was insufficient variation across groups to justify using multilevel analyses (e.g., Thomas & Zhang, 2005). More recently, there has been an increase in the number of published studies using multilevel analysis. Ten of the last 11 issues of the journal Research in Higher Education included at least one study using multilevel analysis (e.g., Bahr, 2012; Cox, McIntosh, Reason, & Terenzini, 2011; Dresel & Rindermann, 2011; Ishitani, 2011; Kim & Sax, 2011; Kugelmass & Ready, 2011; Mayhew, 2012a; Nelson Laird, Garver, & Niskodé-Dossett, 2011; Pike, Kuh, McCormick, Ethington, & Smart, 2011; Webber, 2012). Despite advances in applied statistical techniques that make multilevel analysis
both theoretically and practically preferable to single-level analyses, single-level analyses of multilevel data persist in higher education (Cheslock & Rios-Aguilar, 2008; Zhao & Kuh, 2004). Scholars continue to use OLS regression with multilevel data. (Cole, 2011; Mayhew, Seifert, & Pascarella, 2012).

National Data Set Studies

Over the past 30 years, there has been a rapid acceleration of knowledge creation in higher education. Much of this has come from the evaluation of large and complex data sets. With these data sets, higher education scholars are constructing an increasingly complex and nuanced understanding of the higher education context. The study of the outcomes of college provides an excellent example of this complexity. In the book, *Four Critical Years*, Astin (1977) described the origins of what is now called “college impact.” Prior to the 1960s, few studies considered both intellectual and affective measures as possible college outcomes. In 1965, the University of California at Los Angeles’s Cooperative Institutional Research Program (CIRP) implemented a national and longitudinal study of college students and included both cognitive and affective measures. Data from this study suggested that the college experience was more complex than previously believed and led to the development of Astin’s theory of involvement (1985, 1996).

Later, influenced by Pascarella’s general model for assessing change and the theory of involvement, Astin (1991, 1993) proposed the input-environment-output model (IEO). The IEO model posits that the outcomes of college are a function of both the characteristics and experiences students bring to the college environment. The college
environment consists of institutional characteristics and experiences that act on the student to produce outcomes. Astin’s IEO model is important to higher education research for several reasons. First, the IEO model is general enough that it can be adapted to a wide variety of research topics. Second, the application of the IEO model to college impact research reflects the consensus of scholars that institutional context is significantly associated with multiple outcomes of college attendance. In 2012, the IEO model is a generally accepted framework for studying student outcomes and can be adapted to specific contexts and populations.

There has been consistent growth in the number of studies that analyze multilevel data (Cheslock & Rios-Aguilar, 2008). This growth may reflect the increased prominence of theories that contextualize the college experience. Multi-campus studies allow for analyses based on complex theory, but their sampling methods have the potential to create a problem at the point of analysis. This section presents an overview of major studies that focus on the student experience. Each of these studies has had an important impact in the higher education sphere and has resulted in publications that use hierarchical linear modeling.

**Cooperative Institutional Research Program.** The first sustained national study of college outcomes is CIRP. The CIRP study surveyed students from a national sample of colleges and universities and consisted of collected information about a variety of affective and cognitive outcomes, using both psychological and behavioral measures (Astin, Panos, & Creager, 1967). The CIRP was implemented in response to critiques that research of the era focused too narrowly on academic performance measured by
grades, or graduation rates, and tended to be limited to a single institution. Designed as a longitudinal study, CIRP consists of multiple questionnaires that can be administered to students at various points in their college careers. The study design allows for analysis of data from a single questionnaire or combining multiple questionnaires. Results from the CIRP study form the basis for much of what is known about several decades worth of trends, behaviors, and attitudes about American college students.

**National Center for Educational Statistics.** The National Center for Educational Statistics (NCES), under the direction of the U.S. Department of Education, conducts several national, longitudinal studies related to the college experience. Intended to provide data that support decision making related to national education policy, the NCES surveys include the National Study of Postsecondary Faculty (NSOPF), the Beginning Postsecondary Study (BPS), and the Baccalaureate and Beyond (B&B). Each of these studies is longitudinal, following the same set of individuals over a period of several years, to understand enrollment patterns, faculty mobility, and the economic benefits of completing a college degree (Snyder & Dillow, 2011). A fourth study, the National Postsecondary Student Aid Study (NPSAS), is a combined data set consisting of detailed financial information related to the financial aspects of accessing and completing a college degree. Data from the NCES surveys uses stratified random sample design with oversampling (Cominole, Siegel, Dudley, Roe, & Gillian, 2006). The resulting samples consist of individuals nested within institutions, which indicates that multilevel analysis should be considered when constructing models using these data sets.
National Survey of Student Engagement. The late 1990s brought a time of increased focus on accountability and quality in higher education. In response to these concerns, the Pew Charitable Trust asked several leading higher education scholars to form a panel to explore alternative measures of institutional quality that could be useful to policy makers, campus leaders, and consumers (e.g., student and their families). The result was the creation of a new national study of college students, the National Survey of Student Engagement (National Survey of Student Engagement [NSSE], 2001). The instrument, the College Student Report (CSR), was designed using existing literature on college outcomes (influenced by findings from the CIRP studies) and Chickering and Gamson’s (1987) Seven Principles for Good Practice in Undergraduate Education.

The purpose of NSSE is to “provide data to colleges and universities to assess and improve undergraduate education, inform state accountability and accreditation efforts, and facilitate national and sector benchmarking effort” (NSSE, 2010, p. 7). The first full administration of NSSE occurred in fall 2000 (NSSE, 2001). The study is designed to measure behaviors and attitudes associated with desired college outcomes. These measures are reported as composite measures of engagement, or the extent to which students actively participate in the processes of college. Examples of processes include the number of hours spent studying, using technology for course-related work, interacting with faculty or other higher education professionals. NSSE is modeled as an institutional study used to facilitate institutional improvement efforts and as a national study used to identify experiences, student characteristics, and institutional characteristics associated with engagement.
In addition to the three studies/groups of studies described above, several other national studies have been conducted since 2000. The Multi-Institutional Study of Leadership (MSL) is a study of college student leadership development with an objective to better understand the developmental needs of college students as they relate to leadership and how the college environment affects leadership development (Dugan & Komives, 2007; Multi-Institutional Study of Leadership [MSL], 2009). The Wabash National Study of Liberal Arts Education (The Wabash Study), a longitudinal study of liberal learning outcomes, encompasses a representative sample of institutions and collects qualitative and quantitative data for cognitive and affective outcomes using both affective and behavioral measures (Pascarella & Colleagues, 2007). Finally, there is a growing for-profit industry that licenses/sells rights to use questionnaires with students (e.g., Noel-Levitz). Although infrequently used for scholarly work, these products are another example of how widely surveys are used in the higher education context.

There are several possible explanations for the sudden growth in multi-campus surveys. Led by the work of Astin (1991, 1993), there exists compelling evidence of the need to incorporate institutional context into studies of college outcomes. Failing to consider the influence of institutional context becomes a serious limitation to any study intended to be generalizable. Second, technology makes it possible to administer these studies via online surveys quickly and easily. The number of cases in these studies range from a few hundred to over 100,000 (MSL, 2009; Pryor et al., 2010). Third, advances in both software and hardware make it possible to collect and analyze very large, very complex data sets.
It is likely that higher education scholars will continue to study outcomes embedded in an institutional context or use conceptual models that require the use of predictors measured at different levels. National studies and surveys meet the design needs of this type of investigation. These studies, however, can be costly. As stated previously, the majority of national studies rely on a multilevel sampling strategy. This approach makes it possible to investigate context, but creates problems at the point of analysis when it violates the assumption of independence.

**Statement of the Problem**

Higher education is an applied field, drawing from multiple epistemic and disciplinary perspectives and using a wide variety of methodological approaches (Dressel & Marcus, 1982). Publication patterns in the major higher education journals indicate that higher education researchers focus primarily on describing and understanding phenomena. An unintended consequence observed from these publication patterns is that methodological issues do not become part of the community discourse. Despite a recent public debate about the validity of surveys using cross-sectional designs and self-reported data, this and other issues of method remain largely unresolved (Porter, 2011; Porter, Rumann, & Pontius, 2011).

Prior to the development of multilevel modeling techniques, researchers wishing to model outcomes using multilevel data used ordinary least squares (OLS) regression, knowing that it violated the assumption of independence. Multilevel techniques, including hierarchical linear modeling, take the multilevel data structure into account and provide better estimates of slopes and intercepts. Despite these advances, higher
education has been slow to adopt these techniques. Given the field’s focus on content in its publications, reviews of literature produce little to explain this phenomenon. One possible explanation for the slowness to adopt is that the field, as a whole, does not fully understand how and when to use multilevel techniques.

**Purpose of the Study**

The purpose of this research was to examine how well higher education scholars understand and utilize a specific type of multilevel analysis called hierarchical linear modeling. A search of the literature revealed an unbalanced and incomplete body of knowledge about methodology in general and multilevel analysis in particular. Research on the nature of academic disciplines and discourse communities indicate the literature of the field reflects the values, beliefs, and communal knowledge of a topic (Dressel & Marcus, 1982; Lattuca, 2001). Given the general absence of literature that explicitly addresses methodological issues regarding hierarchical linear models, one strategy to assess higher education’s understanding of this analytic method was to examine how information about the technique is communicated in the studies that report the results of such studies.

In his essay about high quality research manuscripts published in *Research in Higher Education*, Smart (2005) suggested that there are many examples of studies using hierarchical linear models that are applied incorrectly. Smart hypothesized that software makes it easy to perform these types of analyses without fully understanding the statistical theory on which it is based. He also reported that findings to date had not produced results substantively different from those using OLS regression. Smart then
suggested that until scholars were knowledgeable of the methodology, “authors would be best advised to utilize traditional analytic procedures” (p. 467). Thus, the purpose of this research was to examine how well higher education scholars have applied and communicated information regarding methodology in studies that analyzed multilevel data and the reasons presented in publications to justify the selection of single-level (e.g., OLS regression) or multilevel modeling. The focus of this research is on a specific type of multilevel analysis used when the outcome measure is treated as continuous, called hierarchical linear modeling.

**Research Questions**

Based on information in the preceding section and using literature from four major journals of the field, this research is intended to answer the following questions:

1. What methodological sources have been cited by higher education scholars who have published studies that used hierarchical linear models?

2. Using Dedrick et al.’s (2009) analytic framework for examining the narrative content of studies using hierarchical linear modeling, what methodological issues were included/omitted in narratives of studies using the technique?

3. What reasons do scholars give in published articles to justify the use of single-level modeling approaches on complex data?

The answers to these questions provided a baseline understanding of how multilevel data has been modeled in the field and provided insight into perceptions and misconceptions about hierarchical linear models. The findings have implications for the peer-review and publication practices for the journals selected for this study. It also
identified a need to include publications in these journal with a methodological focus. The remainder of this chapter provides additional detail about the context of the research, its significance, and potential contribution to the field of higher education.

**Nature of the Study**

The content analysis research was executed in three parts. In part one, the analysis examined the sources higher education scholars used when citing methodological literature in studies that used hierarchical linear modeling. The second analysis of this research examined the reported content of studies that used hierarchical linear modeling to identify reporting patterns that represent the aspects of methodology that higher education scholars deem relevant to study. The third analysis of this research analyzed the reasons scholars presented to justify the use of single-level modeling techniques (e.g., OLS regression) when analyzing data with a complex structure.

This research applied content analytic techniques to text from empirical studies using hierarchical linear modeling. Data for this dissertation consisted of articles using hierarchical linear modeling from four peer-reviewed higher education journals: *Journal of College Student Development, Journal of Higher Education, Research in Higher Education*, and *Review of Higher Education*. Specific details about how the data set was constructed and analyzed are included in Chapter Three, but briefly summarized here.

Research questions one and two analyzed a data set comprised of studies published since 2000 in the four leading higher education journals that used hierarchical linear modeling. Keyword searches, followed by a systematic review were used to construct the data set. A second sample was created for the analysis of research question
three. Using procedures similar to those used to build the first sample, articles that met the following conditions were included in the study: the data had a nested structure with at least 10 level-2 groups, the dependent measure was continuous, and the data were analyzed using a single-level regression model.

Research question one explored the sources used by authors when writing about hierarchical linear models in methods sections of published studies. Using an approach similar to that of Tight (2008), citations related to methodology were identified and recorded for each study. Citation analysis, a variation of content analysis, identified which citations occurred most frequently in these studies. Results from the citation analysis were reported in tables. Additional analysis examined sources to classify them by type (e.g., book, journal article), and intended purpose within the text.

The analysis for research question two examined how scholars reported the results of hierarchical linear models. Based on a codebook that was adapted from an analysis by Dedrick et al. (2009), this research examined each article in the sample to identify the presence or absence of key elements associated with the methodology of hierarchical linear models. Dedrick et al.’s codebook identified key elements related to the method of hierarchical linear models from primary texts and software manuals about hierarchical linear models. The researchers found that few, if any, studies in their sample included all key elements. This research used Dedrick et al.’s checklist to identify the frequency with which key elements appear in higher education journals. Results were reported quantitatively and qualitatively using charts, tables, and narrative descriptions as is typical of content analyses.
Research question three shifted focus from studies that used hierarchical linear models to studies that analyzed complex data using single-level techniques. Articles meeting inclusion criteria, described in Chapter Three, were analyzed using content analysis to identify what arguments scholars reported, if any, to justify the use of single-level techniques. Because several scholars included a description of a corrective action that produces accurate results, reported corrections were also summarized.

**Significance of the Study**

Results from this research contribute to higher education’s understanding in several ways. The absence of literature about methods in the major journals indicated a need for higher education to attend more carefully to issues of methodology. First, results provide a foundation for understanding how hierarchical linear modeling has been applied to the higher education context. Second, this study extends the work of Dedrick et al. (2009) to include higher education literature and can be used to benchmark the ways in which higher education scholars are similar to and different from the general education field. Analysis of the citations used in reference to methodology of hierarchical linear modeling compiled a comprehensive list of methodological sources for the technique and showed that higher education scholars rely on a relatively small range of sources. Combined, these results made it possible to infer how the field conceptualizes hierarchical linear models and identified gaps in understanding that may explain why, until recently, higher education scholars have been slow to adopt multilevel analyses when working with multilevel data. Finally, there are few higher education focused studies that use the literature of the field as a source of data. This research utilized a
design that can be adapted for other analytic techniques used by higher education scholars. Summaries and syntheses of research methodologies in the field may serve to strengthen the quality of future research.

Definition of Terms

The review of the literature in Chapter Two describes several issues related to the content of multilevel analyses. Some terms, such as “HLM,” have multiple meanings and are defined below for clarification of usage. The following definitions are used in this dissertation.

Multilevel Data

Some sampling strategies create data sets in which groups are selected and then persons are selected from those groups. For higher education researchers, national studies, such as those described here use a similar strategy, with institutions selected first and then students selected from those institutions. The resulting data do not meet the assumption of independent selection, a necessary condition for OLS regression. Multilevel data are also called nested data (Ethington, 1997). For the purpose of this research, multilevel data includes both nested data consisting of variables measured at the same level (nested only) and data sets consisting of variables measured at different levels.

Multilevel Analysis

The term multilevel analysis represents the general class of analytic techniques believed by applied statisticians to be well suited for multilevel data. These techniques do not require data to meet the assumption of independent selection and partition total variance into within- and between-groups components. Different variable types as the
dependent measure are associated with different types of multilevel models. The number of levels of stratification also affects the selection of modeling approach. For example, an outcome variable that is categorical is analyzed using hierarchical multinomial logistic regression (Raudenbush & Bryk, 2002). Multilevel models and multilevel modeling are related terms (Burstein, 1980).

**Multilevel Model**

The term *multilevel model* refers to models that incorporate predictors measured at different levels. For example, an application of Astin’s IEO framework incorporates predictors measured at the student (level-1) and the institution (level-2).

**Hierarchical Linear Models**

The phrase *hierarchical linear model* is used to describe the multilevel model for studies using outcome measures that are continuous or treated as continuous (Gelman & Hill, 2007; Lindley & Smith, 1972; Smith, 1973).

**HLM/HLM6/HLM7**

These terms represent the name of a particular statistical software package used to conduct multilevel analyses on nested data sets. Of all multilevel analysis software packages, HLM seems to be the most commonly used program by higher education researchers to analyze multilevel data. As a result, the term *HLM* is sometimes used to represent the name of a software application, an abbreviation of the phrase hierarchical linear model, or the class of techniques represented by multilevel analysis. For the purpose of this dissertation, HLM is used to represent the software application published by Raudenbush, Bryk, Cheong, Congdon, and du Toit (2011a). It is also used as a
descriptor for one of the samples constructed for this research, the *HLM Studies* sample. In that context, the term represents studies that used hierarchical linear models as an analytic tool.

**Chapter Summary**

This research used the literature of higher education to describe the field’s understanding of the issues related to multilevel modeling. Chapter One included a description of the context and rationale for the proposed study. Chapter Two provides a comprehensive review of the literature related to hierarchical linear models in the higher education literature and includes a summary of the model building process. Chapter Three provides detailed description of the research’s methodology and the limitations associated with the design. Chapter Four summarizes the results of the analyses in this research. Chapter Five includes a discussion of the results from Chapter Four and the implications for higher education research and practice.
CHAPTER TWO
REVIEW OF RELEVANT LITERATURE

This research examined published studies that analyzed data with a multilevel structure and used two data sets: the first consisted of published studies that used hierarchical linear models, the second included published studies that used single-level techniques on multilevel data. The literature reviewed for this research includes publications about the methodology of hierarchical linear models, studies that illustrate the application of the technique in higher education research, and studies that describe the model building process. For this chapter, a search was conducted using multiple search terms in higher education and applied statistics. This produced very few publications that used journal articles as the source of data. The main exception was the literature review and content analysis conducted by Dedrick et al. (2009), which formed the basis for the design for research question two. In addition, this chapter contains a review of literature that expands the context and rationale for the proposed research and describes the conceptual and technical foundations for modeling outcomes using hierarchical linear models.

The chapter begins with a review of literature that forms the rationale for multilevel analysis, including hierarchical linear models. The second section consists of a review of the formulas, terminology and concepts associated with two-level hierarchical linear models and the model building process. The chapter continues with an
examination of the higher education literature related to the methodology of multilevel analysis. The final section of this chapter examines research published in higher education journals that applied a single-level analysis to multilevel data and includes examples of rationales provided to justify the use of single-level models. Studies included in this section provide additional evidence that a need exists to examine the field’s reporting practices related to analyzing multilevel data.

Evolution of Multilevel Analysis

This section is a review of key documents related to the development of multilevel analysis and its introduction to applied fields beginning with a description of the problems associated with modeling multilevel data using single-level techniques, which are both statistical and conceptual. Several sources summarize these issues (Burstein, 1980; Ethington, 1997; Hox, 1995; Raudenbush, 1988; wa Kivilu, 2003).

The Problem of Multilevel Data

Analyzing multilevel data using a single-level method forces the researcher to select one level to use as the basis for analysis. A common data structure in higher education research consists of students clustered within institutions. The data comprise two levels. Data measured at level-1 refer to information about individual students. Level-2 data refer to the institutions. The selection of either level produces increased risk of statistical errors and incorrect interpretation of results (Robinson, 1950; Simpson, 1951).

Empirical evidence. Robinson (1950) was one of the earliest to publish on the problem of cross-level inferences. He observed that many sociological studies included
statistical analyses at the group level, using the term *ecological* to represent the group. Robinson argued that the choice of the group as the unit of analysis was due to difficulty acquiring data at the individual level rather than an intentional choice to draw inferences about the group level. He analyzed data from the 1930 U.S. Census and showed statistical relationships, measured using correlations, at one level did not hold for the other level. Robinson concluded that the analysis “provides a definite answer as to whether ecological correlations can validly be used as substitutes for individual correlation. They cannot” (p. 357). Reflecting the language used in Robinson’s article, a result of this type is called an *ecological fallacy*, or the Robinson effect. Reducing the risk of ecological fallacy is a frequently cited rationale for using multilevel analysis (Patrick, 2001; Porter, 2006; Porter & Umbach, 2001; Smeby & Try, 2005).

Cronbach and Webb (1975) provided additional evidence against the use of single-level analyses with multilevel data. They reanalyzed the data from a previous study testing the effectiveness of different teaching methods on student achievement. The original study found an interaction effect between ability, measured by the Minnesota Ability Test, and treatment, which consisted of two approaches to teaching arithmetic (Anderson, 1942). Cronbach and Webb reanalyzed data to “separate between-class and within-class components” (p. 717) and found that the interaction effect was no longer significant. The authors interpreted their findings in the context of assumptions made by scholars regarding differential effects of instruction, but cautioned against generalizing their results. For the purposes of the present review, the results serve as empirical evidence, despite some suggestions to the contrary, that single-level and
multilevel analyses do not persistently produce similar findings. This study is therefore important in the context of higher education research as some scholars in that field incorrectly argue that OLS regression on multilevel data provide essentially the same results as multilevel techniques (Astin & Denson, 2009).

Burstein, Linn, and Capell (1978) conducted a study that compared and contrasted single-level and multilevel modeling on an academic outcome. They used simulated data that grouped students within classrooms and had significantly different means scores on the outcomes across groups. Given prior knowledge regarding the variation between groups, Burstein et al. were able to test several hypotheses including: “[a]nalyses conducted at only a single . . . level are usually inadequate” and “[w]hen within-class regression functions differ and the analyst treats them as equal . . . much of the substantive information about within-class and between-class effects can be masked” (p. 349). They constructed two single-level models, at the student and class levels, and tested three multilevel approaches. Their results confirmed their hypotheses and they concluded that one method, slopes-as-outcomes, produced the most accurate estimates. Supporting the results of Cronbach and Webb (1975), Burstein et al. presented empirical evidence that single-level analyses were not sufficiently robust to use with multilevel data.

**Statistical assumptions.** In addition to empirical evidence regarding the problems of applying single-level analyses to multilevel data, there are several issues that arise due to violations of statistical assumptions. Higher education researchers have emphasized how violating the assumption of independence increased the probability of Type I error
due to underestimated standard errors (Ethington, 1997; Thomas & Heck, 2001). Hox and Kreft (1994) described several problems related to aggregating or disaggregating multilevel data so it can be analyzed at a single-level. They used the perspectives of sociology and ANOVA to frame their argument and identified several problems related to single-level analyses of multilevel data.

The first issue described by Hox and Kreft (1994) was the assumption that single-level analyses require errors to be independent. Applied to the two-level data structure described previously, the assumption of independence means that there is no statistical relationship between students in a group on the outcome measures (Tabachnick & Fidell, 2001). Hox and Kreft suggested that the magnitude of correlation between error terms can be assessed using the intraclass correlation (ICC) and cited a study by Barcikowski (1981) that showed an ICC of .01 for a balanced sample of 100 observations across four groups increased the Type I error from .05 to .17. The ICC is a statistic that describes the proportion of total variance that can be attributed to differences between groups (Raudenbush & Bryk, 2002).

Another issue identified by Hox and Kreft (1994) related to the nesting that is typical of multilevel data. The variance of nested data consists of two parts: a within-group component, and a between-group component. Traditional ANOVA can partition this variance in balanced designs, but not so efficiently with unbalanced data. Hox and Kreft argued that multilevel models produce unbiased estimates for unbalanced data making the technique preferable to single-level approaches. A third issue is related to modeling interaction effects. An interaction effect occurs “when the effect of one
independent variable on the dependent variable depends on the level of the second independent variable” (Pallant, 2007, p. 257). A cross-level effect is an interaction effect in which the predictors that vary jointly are from different levels. Hox and Kreft showed that modeling cross-level interactions on multilevel data at the lowest level (i.e., using students as the unit of analysis) further biases estimates and standard errors.

**Statistical Evidence**

While applied researchers built a body of empirical evidence against single-level analyses, statisticians examined the issue from the point of view of statistical theory. Prior to reviewing the literature of this section, it is helpful to review the statistical form of a two-level hierarchical linear model. Equations (1) and (2) represent each level of the two-level model

**Level-1**

\[ Y_{ij} = \beta_{0j} + \beta_{1j}(X_{1ij}) + \beta_{2j}(X_{2ij}) + ... + \beta_{nj}(X_{nij}) + r_{ij} \]  

(1)

**Level-2**

\[ \beta_{ij} = \gamma_{0j} + \gamma_{1j}(Z_{1j}) + \gamma_{2j}(Z_{2j}) + ... + \gamma_{nj}(X_{nj}) + u_{ij} \]  

(2)

\( Y_{ij} \) represents the value of the outcome measure for person \( i \) in group \( j \). \( \beta_{0j} \) represents the intercept for group \( j \). Each \( \beta_{nj} \) represents the slope (effect) for variable \( X_{nj} \) on group \( j \). The \( X \) terms represent level-1 independent variables and the \( Z \) terms represent level-2 independent variables. The individual error terms \( r_{ij} \) are assumed to be normally distributed with a mean of 0 and variance \( \sigma^2 \). Conceptually, this looks similar to two groups of regression equations. However, the total variance is now written as two components, a within-group component (\( r_{ij} \)) and a between-group component (\( u_{ij} \)). Given the similarity to OLS regression, it may be tempting to argue OLS algorithms can be applied to the data to construct a multilevel model. However, citing a previous study,
Hox and Kreft (1994) contend:

Using ordinary multiple regression to estimate the regression coefficients in both steps is inconsistent because, in the first step we view the regression coefficients $\beta_p$ as fixed coefficients to be estimated by within-groups regressions, whereas in the second step, we view them as random variables to be estimated by a between-groups regression. If the OLS assumptions are true at the individual level, they will not be true at the second; the error structure will generally be quite different from the error structure assumed by the linear model. As a result, significance tests based on the usual standard errors are badly misleading. (p. 288)

Traditional OLS regression procedures do not decompose the variance, which leads to increased Type I error due to underinflated standard error estimates (Raudenbush, 1995; Raudenbush & Bryk, 2002).

**Developments in statistical theory and numerical procedures.** A detailed examination of the theory around multilevel analysis requires an understanding of advanced statistical theory. In addition to the literature reviewed in this section, there are several publications that contributed to the development of multilevel analyses. This literature is summarized in Table 1. Selected articles are summarized and discussed in the section that follows the table.

**Table 1. Contributions to the Statistical Development of Multilevel Analysis**

<table>
<thead>
<tr>
<th>Author/Year</th>
<th>Title</th>
<th>Type of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lindley and Smith (1972)</td>
<td>Bayes Estimates for the Linear Model</td>
<td>Theory</td>
</tr>
<tr>
<td>Smith (1973)</td>
<td>A General Bayesian Linear Model</td>
<td>Theory</td>
</tr>
<tr>
<td>Dempster, Laird, and Rubin (1977)</td>
<td>Maximum Likelihood From Incomplete Data Via the EM Algorithm</td>
<td>Numerical</td>
</tr>
<tr>
<td>Burstein (1978)</td>
<td>Assessing Differences Between Grouped and Individual-level regression coefficients</td>
<td>Theory</td>
</tr>
</tbody>
</table>
Lindley and Smith (1972; Smith, 1973) established the theoretical foundation for contemporary hierarchical linear models. Their studies established that Bayesian estimates are applicable to linear models for multilevel data. These models include a variance structure that separates total variance into between- and within-group components. Several reviews acknowledge this contribution, but note that numerical approaches did not develop in parallel (Ethington, 1997; Raudenbush & Bryk, 2002).

The expectation maximization (EM) algorithm generates maximum likelihood parameter estimates. Dempster, Laird, and Rubin (1977) showed that the EM algorithm produces accurate estimates for parameters. There are two types of maximum likelihood estimates, full (ML) and restricted (REML). These are iterative procedures that require convergence to report estimates. Harville (1977) compared the results of ML and REML
algorithms and found that REML corrects for bias to estimates that can occur in ML procedures. Full information maximum likelihood (FIML) is similar to ML and REML and is used in some software applications to generate parameter estimates.

Because there are several theoretically correct numerical approaches and the fact that ML and REML algorithms produce different results, it is important to know which algorithm is used when conducting multilevel analyses. De Leeuw and Kreft (2011) summarized stand-alone software and modules intended to conduct multilevel analysis including HLM and PROC MIXED (Raudenbush et al., 2011a; Statistical Analysis System [SAS], 1992). De Leeuw and Kreft did not identify a superior algorithm to use in multilevel analyses. They did, however, report that REML is the default for two-level models and FIML is the default for the three-level model in HLM7. Users may adjust the software settings to select a different algorithm.

**Applying Multilevel Analysis to Education Research**

Ethington (1997) authored what may be the first publication that made an explicit connection between multilevel analyses and higher education research. Ethington argued that there were statistical and conceptual arguments to justify the use of hierarchical linear models. Citing Pascarella’s general model for assessing change (1985) and Weidman’s model of undergraduate socialization (1989) as examples, Ethington argued that contemporary models of college impact included both student and institution level characteristics and that these frameworks made it necessary to collect data about students and the institutions they attend. The resulting data set has the multilevel structure that forces aggregation or disaggregation of data in order to conduct single-level analyses.
Ethington then summarized the conceptual bases for hierarchical linear models and presented the results of a study using CIRP data to model educational attainment. The concepts of hierarchical linear models are presented in a later section of this chapter.

In a complementary publication to Ethington’s (1997) description of the application of hierarchical linear models to higher education data sets, Thomas and Heck (2001) wrote about the challenges of analyzing “large scale secondary data” (p. 517). In a widely cited article, *Analysis of Large-Scale Secondary Data in Higher Education Research: Potential Perils Associated with Complex Sampling Designs*, Thomas and Heck reported an increase in the use of large data sets, complex sampling designs in higher education research, and identified two issues that can lead to statistically incorrect results if not addressed correctly when analyzing the data.

The first issue described was that sampling strategies for large scale data sets frequently oversample some subgroups to increase the likelihood that a sufficient number of subgroup members were included in the final data set. The solution to the problem of oversampling is to use weighting procedures to statistically rebalance the data. Weights are frequently included as a variable in data sets used for secondary data analysis and should be included in both single-level and multilevel analyses (Thomas & Heck, 2001).

The second issue related to these data sets was that multistage sampling designs produce multilevel data. As has been stated previously, this violated the assumption of independence. Failure to address the statistical effect of clustered data produces estimates that are biased because traditional software applications analyze data at a single-level. The issue of multilevel data does not have a single solution according to
Thomas and Heck (2001). They described two classes of adjustments that can be made to produce unbiased estimates: model-based or design-based. Modeling the data using a multilevel analysis is considered a model-based approach and is the only solution of this type mentioned by Thomas and Heck. Model-based approaches are the focus of this research and were the preferred solution according to Thomas and Heck. The findings of this study suggest that scholars, instead of selecting the recommended approach, adopted practices that were identified as less preferable by Thomas and Heck, such as using more conservative $p$-values in hypothesis tests.

Design-based approaches are adjustments made to data or the statistical analysis to correct for biased estimates. The first design-based approach consists of specialized procedures in statistical software applications. Thomas and Heck (2001) identified several stand-alone software applications that use bootstrapping, Taylor expansion, and jack-knifing techniques to produce unbiased standard errors, but also noted that at the time the software was not simple to use. In their analysis, Thomas and Heck compared the results of a weighted OLS regression analysis and an adjusted model. They found that the parameter estimates were identical to three decimal places but that standard errors were inflated in the OLS regression, leading to different results from hypothesis tests.

The second type of correction presented by Thomas and Heck (2001) was the recommendation to use the root mean design effect (DEFT) in hypothesis tests. Standard errors are multiplied by the DEFT to produce adjusted standard errors. Thomas and Heck commented that the DEFT is not always easily calculated, but that NCES data sets tend to
include them in their data. The findings from this research showed that scholars did not select this option as a corrective measure for analyzing multilevel data. The third adjustment proposed by Thomas and Heck was to “alter the effective sample size by adjusting the relative weight downward” (p. 533). Not all software applications allow for this solution. Finally, Thomas and Heck suggested using a more conservative $p$-value in hypothesis tests. This recommendation was found to be the most frequently applied corrective measure by higher education scholars.

One of the reasons Thomas and Heck (2001) is cited frequently in the higher education literature is because it offered specific guidance on how to adjust for non-independence. It is worth noting, however, that despite the fact that Thomas and Heck stated explicitly that conducting a multilevel analysis is the preferred method for dealing with non-independence, several scholars select the least recommended correction: use of a conservative $p$-value. Hypothesis tests are sensitive to sample size, and studies using the types of data sets described here can consist of several thousand cases. It may be that the traditional “conservative” $p$-value, typically less than .01, is not conservative enough.

**Building and Reporting Hierarchical Linear Models**

The review of literature produced few studies that included recommendations for the content of reporting hierarchical linear modeling (Dedrick et al., 2009; Ferron et al., 2008; McCoach, 2010). In their review of the technical literature related to hierarchical linear models in the general education literature, Dedrick et al. (2009) identified four themes related to methodology of hierarchical linear models. In a book chapter based, in part, on the study that informed Dedrick et al., Ferron et al. (2008) identified those items
related to model developments and specification that should be included in reporting of hierarchical linear modeling studies. This content analysis study extends these results in two areas by focusing on higher education journals and including publications from 2000 to 2012.

The Sample

Good scholarship includes a detailed description of the sample used for the analysis. This information includes a description of data collection procedures, including sampling strategies, and the characteristics and size of the sample (Dedrick et al., 2009; Ferron et al., 2008). There is no consistent pattern of sampling strategies for higher education research studies using hierarchical linear models.

The National Center for Educational Statistics (NCES) uses a traditional complex sample design in which random sampling procedures are used to identify individuals at level-1 and institutions at level-2. The Higher Education Research Institute (HERI) uses nonrandom samples at both level-1 and level-2 in combination with weights to make the data representative of all four-year colleges and universities (Higher Education Research Institute, n.d.). Other studies used a mixed sampling approach in which the level-2 sample is not random, but the level-1 sample is random. The Multi-Institutional Study of Leadership, for example, uses mixed samples at the majority of participating campuses (Multi-Institutional Study of Leadership [MSL], n.d.). These sampling strategies are qualitatively different from those found in other literature bases. In the general education context, for example, a more commonly applied mixed sample uses a random sample at level-2 and a non-random sample at level-1. In this design, intact classrooms are
randomly sampled for K-12 research. The resulting data are hierarchical in structure with a random sample at level-2 and a non-random sample at level-1. Sampling strategies affect hierarchical linear modeling studies in several aspects. The majority of these studies use sample weights. Reporting results should indicate when weights are included in the analysis and the type of weighting if multiple weights are available.

Information about the sample should be reported at each level of the model. A balanced sample will have the same number of cases in each level-1 group. Sample size should be reported as the total number of level-1 cases, the number of level-2 units and the number of cases in each level-1 unit (McCoach, 2010). Unbalanced samples can be represented in different ways. Ferron et al. (2008) recommended reporting on the total number of level-2 groups. If the number of level-2 groups is greater than 50, then report the following: (a) the total number of level-1 cases, (b) the total number of level-2 cases, (c) the range of level-1 sample sizes, and (d) the mean and standard deviation of the level-1 sample sizes. This allows the reader to understand the range of level-1 sample sizes. For samples with fewer than 50 level-2 groups, Ferron et al. recommended representing the level-1 sample sizes using tables or stem-and-leaf plots.

The size of the level-2 and level-1 sample influences the modeling process. Estimation procedures, described later, produce parameter estimates with decreasing bias as the sample size increases. Because hierarchical linear modeling produces parameter and variance estimates, the number of level 2 groups required varies. Maas and Hox (2005) conducted a simulation study in which sample size at each level and ICC were varied to determine the minimum size at each level to assure unbiased estimates. They
tested all permutations of the following conditions: level-2 sample size (30, 50, 100); level-1 within-groups sample size (5, 30, 50); and ICCs (.1, .2, .3). The values selected for each condition reflected typical sample characteristics in educational and organizational research. Maas and Hox found that if the coefficients were the primary focus of the study, then a level-2 sample may be as small as 10 groups. The standard errors of the level-2 variance components were unbiased only under the condition of 100 level-2 groups. Applications of hierarchical linear modeling in higher education use group sizes ranging from small (fewer than 20 level-2 groups) to very large (greater than 500).

Examining the power of a study should be a consideration of any research that uses hypothesis testing. Hypothesis testing examines uses probabilities to determine whether the extent of a difference or relationship between two values is likely to occur at random. Power is a measure of the likelihood that one would accept the alternative hypothesis when, in reality the null hypothesis is correct (Cohen, 1988; Dattalo, 2008). Power is influenced by multiple characteristics of a research design including: sample size, the statistical test selected for the analysis, the intended or actual effect size, and probability of Type I error (α) (Murnane & Willett, 2011). McCoach (2010) and Dedrick et al. (2009) both suggest that power should be reported in studies using multilevel modeling. Dedrick et al. found that few studies in their sample of general education literature reported statistical power. Preliminary scans of the literature for this research appear consistent with this finding.
Murnane and Willett (2011) claim that power can be increased through the use of more complex models, arguing that increasing the number of variables in the model adds precision but also caution that use of more complex techniques requires understanding of the implications of the assumptions of advanced statistical techniques. Failure to attend to the statistical details creates more opportunity for error. The more highly correlated the outcome variable is within a level-2 group the greater the reduction in statistical power. In a simulation, Murnane and Willett modeled the relationship between power, the number of level-2 groups, and the intraclass correlation. In their simulation one might need as many as 75 level-2 groups to maintain a minimal power of .8 when the ICC is 10%. It may be tempting to use the ICC as an estimate of statistical power. However, as Murnane and Willett point out, increasing sample size may increase statistical power. It may also, however, increase the intraclass correlation which reduces power. As such, several who have written on statistical power recommend calculating statistical power for these types of studies. Because the relationship between statistical power and characteristics of a research design are not linear statistical power must be established numerically. The complexity of the formula for multilevel models makes the use of software preferable to manual approaches (c.f., Murnane & Willett, 2011). Both McCoach and Dedrick et al. cite Optimal Design software as a tool for conducting a priori power analyses (Raudenbush et al., 2011).

Model Development and Specification

This section consists of a synthesis and review of the recommended structure and format for reporting results of hierarchical linear modeling studies. The first step of
model specification is to present a justification for the use of hierarchical linear models. Dedrick et al. (2009) found that the majority of studies ($n = 86$) in their sample provided a rationale for the use of hierarchical linear models, but that the level of justification varied. Although not noted explicitly by any of the authors cited here, the justification and rationale should indicate the precise analytic method. Hierarchical linear models are appropriate for multilevel modeling of outcomes that are continuous and normally distributed. Hierarchical linear regression refers to a single-level analysis and hierarchical generalized linear models model outcomes that are not normally distributed. Failure to use consistent and specific language may lead to confusion particularly in those studies that are applying a single-level analysis to multilevel data.

**Statistical models.** This research examined empirical studies that analyzed data with a multilevel structure. These studies model multilevel data with outcomes that are continuous and normally distributed. There are different approaches to describing these models mathematically (Hox, 1995, 1998; Raudenbush & Bryk, 2002; Singer, 1998). Both McCoach (2010) and Ferron et al. (2008) indicated that the statistical model should be presented in the methods section of a multilevel analysis. Dedrick et al (2009) did not report the frequency with which statistical models were included in the sample. McCoach suggested that some journals do not want an explicit statement of the statistical model based on readership. The submission guidelines for *Journal of Higher Education*, for example, recommends that methodological detail be brief, as the emphasis of the narrative should be on content findings and implications.
The statistical model for a two-level hierarchical linear model can be written in one of three forms. The first form consists of a system of equations. Using a simple model with a single predictor at each level, $X_1$ and $Z_1$, respectively, equation (3) represents the level-1 model. It assumes the form

\[
\text{Level-1} \quad Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij}) + r_{ij} \quad r_{ij} \sim N(0,\sigma^2)
\]  

$Y_{ij}$ represents the value of the outcome measure for person $i$ in group $j$. $\beta_{0j}$ represents the intercept for group $j$ and $\beta_{1j}$ represent the regression slope for variable $X_{ij}$. The individual error terms $r_{ij}$ are assumed to be normally distributed with a mean of 0 and variance $\sigma^2$. This model appears similar to an OLS regression equation with a key difference. The intercepts and slopes can be modeled using level-2 predictors. Equations (4) and (5) represent the modeling of the Level 1 parameter estimates.

\[
\text{Level-2} \quad \beta_{0j} = \gamma_{00} + \gamma_{01}(Z_j) + u_{0j}
\]

\[
\beta_{1j} = \gamma_{10} + \gamma_{11}(Z_j) + u_{1j}
\]

Again, the level-2 models appear similar to an OLS regression equation. Each level-2 equation includes a separate error term, $u_{0j}$ and $u_{1j}$. The level-2 error terms are assumed to have a multivariate normal distribution and are independent of the level-1 error, $r_{ij}$.

The statistical model can also be represented as a single multilevel model of the form:

\[
Y_{ij} = [\gamma_{00} + \gamma_{01}(Z_j) + u_{0j}] + [\gamma_{10} + \gamma_{11}(Z_j) + u_{1j}](X_{ij}) + r_{ij}
\]

\[
= \gamma_{00} + \gamma_{01}(Z_j) + \gamma_{10}(X_{ij}) + \gamma_{11}(X_{ij})(Z_j) + u_{ij}(X_{ij}) + u_{0j} + r_{ij}
\]  

McCoach (2010) made specific mention of the need to indicate which variables are allowed to randomly vary across groups. A level-2 predictor that does not vary randomly
across groups will be fixed for all groups and the level-2 equation will not include an error term. This form was reported less frequently in higher education literature, but may aid in the interpretation of results as software packages such because HLM 7 report results for the $\gamma$s. The third form used to represent the statistical model uses matrix notation and was not used in any of the articles included in this research.

Several questions emerge from the review of the literature regarding presentation of the statistical model. First, is a general form of the model sufficient to meet the recommendation or should the statistical model reflect the variables used in the actual analysis? Second, should the statistical model reflect the final model of the data and include only those variables that are significant? Finally, what are the possible implications of not explicitly identifying which effects are fixed and which are allowed to vary randomly across groups? This last issue may be of greatest importance as one of the key differences between single-level regression and multilevel modeling is the partitioning of variance into between- and within-group components, and to allow some variables to not vary across groups. Failure to report this may contribute to some of the misconceptions that multilevel analysis is equivalent to OLS regression among those not familiar with multilevel techniques.

**Variable selection.** The identification of variables to include in the model is an integral part of modeling outcomes. McCoach (2010) suggested that variable selection for a model be consistent with the research questions and based on theory. McCoach acknowledged a temptation to include any variable that may be significant given that excluding “an important potential confounder creates the potential for bias in the
estimates” (p. 125), but cautioned against including too many variables, which may produce models that are difficult to interpret. McCoach cites Raudenbush and Bryk (2002) as justification for a recommendation to not allow all level-1 variables to vary randomly. Ferron et al. (2008) and Dedrick et al. (2009) did not express the same level of caution regarding variable selection related to fixed and random effects. Ferron et al. noted only that each selected variable should be classified as an outcome, predictor, or covariate and Dedrick et al. did not include this level of detail in their analysis. There was consensus across the literature that descriptive statistics for all variables tested in the model should be presented and should include information about coding and transformations. Both McCoach and Ferron et al. mentioned the need to verify and report reliability tests for composite measures. Ferron et al. were specific in their guidance. If the sample used in a reported study is a subsample of a larger data set evidence of reliability must be based on the sample for the study, not the larger data set. Low reliability increases the likelihood of Type II errors.

**Centering.** Based on the literature, linear transformations of variables are a common practice in hierarchical linear modeling studies. Variables are often centered to make it easier to interpret parameter estimates. There are three types of linear transformation strategies to center the data. The first, grand mean centering, subtracts the grand mean of a variable X from $X_{ij}$. Ferron et al. (2008) reported that is statistically appropriate for both level-1 and level-2 variables. The second strategy, group mean centering, subtracts the group mean from the variable $X_{ij}$. Ferron et al. claimed this transformation should not be used with level-2 variables. Dedrick et al. (2009) and
Ferron et al. suggested the primary reason for centering is to correct the metrics of a continuous variable and offer the example of a standardized test score with a range from 200 through 800. Interpretation of coefficients assumes the value of the predictor is set to zero, which is inconsistent with measures that never assume the value zero.

**Variance structures.** One of the key features of hierarchical linear models is that the statistical models include multiple error terms. The existence of a more complex covariance structure makes it possible to allow for the inclusion of different types of research questions related to differences across groups (Dedrick et al., 2009). Bryk and Raudenbush (2002) stated that the default assumption is that all errors are homogeneous at both level-1 and level-2 and that this will apply to a majority of models. Complex level-2 equations, which model level-1 slopes, include a combination of fixed and randomly varying predictors. The many possible combinations of level-2 equations can make it difficult to ascertain when covariance structures are not defined appropriately.

Testing model fit is another element that should be included when reporting the results of hierarchical linear models (McCoach, 2010). Singer (1998) suggested that indices such as the Akaike Information Criterion (AIC, Akaike, 1974) and the Bayesian Information Criterion (BIC, Schwarz, 1978) can discriminate among models to identify the one with the “best fit.” Each index produces a number based on the log likelihood and the number of estimated parameters. The closer the value of the criterion, the better the model fits the data.
Data Preparation

Similar to single-level regression analyses, data should be prescreened to identify potential sources of bias and error. One of the rationales for multilevel analysis is that it does not require the data to be independent. Multilevel analyses does require, however, that the data and errors be normally distributed and that the variances are homogeneous at both levels (Dedrick et al., 2009; Raudenbush & Bryk, 2002; Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011b). Dedrick et al. (2009) suggested that one of the most important data related issues to address is to verify that the residuals are normally distributed and the variances are homogeneous. The literature included in this review recommended these assumptions be tested. Dedrick et al. recommended that normality can be tested using the distribution of residuals at each level. Taking the natural log of a variable that is not normally distributed is an accepted correction (Tabachnick & Fidell, 2001).

Outliers and missing data. A second source of bias in statistical analysis comes from the presence of outliers. Outliers in hierarchical linear models contribute to distortion of results in models that are similar to effects found in regression modeling. Outliers can be identified using univariate statistical analyses including stem-and-whisker plots. Missing data can also affect results of hierarchical linear modeling analyses. Bias is introduced into models if too much data are missing and if the data are not missing at random. To test whether or not the data are missing at random, Tabachnick and Fidell (2001) recommended testing for significant differences in the mean scores for groups created when separated into missing on variable X and not missing on variable X. If
there are no significant differences, then the missing data can be treated as “at random.”

The corrections for missing at random data include dropping cases with missing data, omitting the variable from the analysis, or estimating a value to use in place of missing data. There is consensus across sources that if data are missing, the reporting of the study should include text that describes the extent to which the data are missing, whether or not it is missing at random, and procedures for managing the missing data (Dedrick et al., 2009; Ferron et al., 2008; McCoach, 2010).

**Measurement error.** The final data consideration is to examine measurement error. All data have some level of measurement error (Dedrick et al., 2009). Reporting of results should include a description of how the statistical reliability of measures was established. In the case of composite measures, a frequently used measure in the higher education literature, reliability is calculated using the actual sample for the study. Reliability measures calculated on the full data set may serve as a guide, but are not sufficient to establish accurate measurement in an individual study (Thompson & Vacha-Haase, 2000).

**Estimation Techniques**

One of the key differences between traditional OLS regression and hierarchical linear modeling is the method of estimation. Estimation methods are theoretical approaches to estimating the fixed effects, random coefficient, and the variance and covariance components (Raudenbush & Bryk, 2002). The most common estimation methods used in hierarchical linear models are maximum likelihood (ML), restricted maximum likelihood (REML), and Bayesian estimates (Ferron et al., 2008). There are
different numerical methods, or algorithms, that can be applied for each estimation method. Both ML and REML estimation seeks to determine the parameter estimate that “maximize the likelihood of the data” (Dedrick et al., 2009, p. 80). This is achieved using an iterative algorithm until the model converges, or reaches a point where the difference between estimates reaches a predefined minimum value.

Maximum likelihood estimates can be conducted using the expectation maximization (EM) algorithm, the Newton-Raphson algorithm or the Fisher scoring algorithm (Dedrick et al., 2009). ML uses the same algorithm for both the parameter and variance estimates. In contrast, REML, uses the ML algorithm to estimate the variance parameters and generalized least squares algorithm to estimate fixed effects. Bayesian estimation, based on the work of Lindley and Smith (1972; Smith, 1973) uses Markov Chain Monte Carlo algorithm. HLM7, a commonly used software application in education research, uses REML for hierarchical linear models and a combination of Fisher scoring and generalized least squares to estimate fixed effects and variance and covariance parameters (Raudenbush et al., 2011b).

Both McCoach (2010) and Dedrick et al. (2009) commented that there is no estimation method that works in all cases, which is likely one of the rationales for including information about estimation methods when reporting results of hierarchical linear models. For data sets consisting of large numbers of level-2 groups, REML and ML produce similar estimates of the variance components. For smaller sized level-2 samples, REML is preferred to ML because ML variances are underestimated (McCoach, 2010). In their study of general education journals, Dedrick et al. found that 15 of 98
articles reported information about estimation method and information about algorithms used for estimation was reported in 0 out of 98 articles. The low incidence of reporting information about estimation methods and algorithms may reflect journal guidelines that emphasize content over methodological detail but may also reflect an author’s lack of knowledge regarding the effect estimation can have on parameter and variance estimates.

**Model Building**

A two-level hierarchical linear model is characterized by the presence of predictors measured at two levels that also partitions the variance into between- and within-group components. Modeling of this type in higher education research frequently consists of students who are members of an institutional “group.” In the language of multilevel analysis, students are referred to as level-1 units and the institutions are at level-2. Level-1 predictors refer to variables collected about an individual student. Level-2 predictors reflect institutional characteristics such as institutional type (e.g., public/private) or an aggregated measure such as the percent of first generation students.

Two types of means are included in two-level hierarchical linear models. First is a grand mean, the mean for all individuals in the sample. The second mean is the group mean, or the mean of all members within a particular group. Independent variables are represented by X at level-1 and Z at level-2. Finally, the purpose of multilevel analysis is to separate variance into between- and within-group components. Each of the models presented below includes error components at both the student and institutional level.

**Fully unconditional model.** The first step to conducting a multilevel analysis consists of conducting an analysis at the group level and tests for significant differences
in the group means. This model is called the *fully unconditional model*. The statistical form for the fully unconditional model is represented in equations (7) and (8).

\[
\text{Level 1: } Y_{ij} = \beta_{0j} + r_{ij} \quad \text{where } r_{ij} \sim N(0, \sigma^2)
\]

\[
\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j} \quad \text{where } u_{0j} \sim N(0, \tau_{00})
\]

$Y_{ij}$ is the value of the outcome measure for student $i$ at institution $j$. $\beta_{0j}$ is the value of the intercept for institution $j$, or the mean outcome for the $j$th group. $\gamma_{00}$ represents the grand mean. The level-1 error term, $r_{ij}$, represents a random error associated with student $I$ at institution $j$. The level-2 error term, $u_{0j}$, represents the error associated with the group means. Both the level-1 and level-2 errors are assumed to be independent and normally distributed with means of 0 and variances $\sigma^2$ and $\tau_{00}$, respectively. The hypothesis test for the model tests whether or not the variance components are significantly different from zero.

The fully unconditional model is conceptually equivalent to a one-way ANOVA with random effects (Raudenbush & Bryk, 2002). A significant result in the fully unconditional model means that at least one of the groups is significantly different than the others on the outcome measure. One of the most common ways that higher education scholars report the results of this model is by reporting the ICC (Gelman & Hill, 2007; Raudenbush & Bryk, 2002; Singer, 1998). This statistic describes the proportion of total variance in the outcome measures that can be attributed to differences in the group means. The formula for the ICC is $\rho = \tau_{00} / (\tau_{00} + \sigma^2)$ where $\tau_{00}$ represents the between-group variability and $\tau_{00} + \sigma^2$ represents the sum of the between- and within-group variance, or the total variance.
McCoach (2010) observed that the purpose of estimating the fully unconditional model is to “obtain estimates of the level-1 and level-2 variance components for comparison to later…models and to estimate the ICC” (p. 135) and stated explicitly that the ICC should be reported and interpreted. Studies from the early 2000s, however, reported the ICC less than 50% of the time (Dedrick et al., 2009). One possible explanation for the finding may be that the authors wanted to frame results as being similar to OLS regression or space limitations set by journals. A preliminary review of studies from higher education journals suggests that when the fully unconditional model is reported, the emphasis is on reporting the ICC as a justification for using hierarchical linear modeling. The results of this research offer insight into when and how the fully unconditional model is used in the higher education context.

**Random coefficients model.** Similar to model outcomes using a single-level regression analysis, the process of building a completed hierarchical linear model may include several preliminary models using different combinations of level-1 and level-2 predictors. Dedrick et al. (2009) found that the number of models tested in studies included in their analysis ranged from 1 to 430. The majority of articles in their study reported using between 1 and 10 models. McCoach (2010) stated that a specific type of hierarchical linear model, the *random coefficients model*, should be reported and include a table that includes “both the fixed effect parameter estimates and variance components” (p. 124). Dedrick et al. did not report specific types of hierarchical linear model structures, focusing more on the types and combinations of predictors in the analysis (e.g., interaction effects, cross-level effects).
The random coefficients model includes level-1 predictors similar to what one would find in an OLS regression using level-1 as the unit of analysis. There are two differences in the random coefficients model compared to OLS regression. First, only level-1 variables are included in the model, which excludes any level-2 variables. This is a reflection of the conceptual foundation of multilevel analysis, that the level of the variable must match the level at which it is entered. Second the parameter estimates are identical for all cases in the data set. The random coefficients model allows the level-1 parameter estimates (slopes) to vary randomly across groups (Raudenbush & Bryk, 2002). A random coefficient model with two level-1 predictors is represented using the equations (9) through (12).

\[
\text{level-1:} \quad Y_{ij} = \beta_{0j} + \beta_{1j}(X_{1ij}) + \beta_{2j}(X_{2ij}) + r_{ij} \quad \text{where } r_{ij} \sim N(0, \sigma^2) \quad (9)
\]

\[
\text{level-2:} \quad \beta_{0j} = \gamma_{00} + u_0 \quad (10)
\]
\[
\beta_{1j} = \gamma_{10} + u_{1j} \quad (11)
\]
\[
\beta_{2j} = \gamma_{20} + u_{2j} \quad (12)
\]

The presence of the level-2 error terms represents, statistically, that the intercept, \( \beta_{0j} \), and the slopes, \( \beta_{1j} \) and \( \beta_{2j} \), can vary randomly across the level-2 groups, but that this variation is not modeled using any level-2 predictors (Raudenbush & Bryk, 2002). The level-2 variance and covariances are called unconditional because there are no level-2 predictors. It is also important to note that level-1 predictors are frequently, but not always centered in a hierarchical linear model. Centering is not represented in the example of the statistical model for a random coefficients model.

The results of the random coefficients model includes parameter estimates for...
each of the $\gamma$s in the model and the variance for each of the error terms. The parameter estimates are called fixed effects and are tested for significance using a $t$-ratio. A significant fixed effect should be included in later models. The variance estimates are called random effects and are tested using a $\chi^2$ statistic. If a variance estimate is significant, then there exist differences across groups for that parameter and the variable should be allowed to randomly vary in later models.

**The full model.** A full hierarchical linear model consists of modeling the outcome using both level-1 and level-2 predictors. The results of the random coefficients model provide two important results that are used in the development of the full model. First, the results of the random coefficients model are used to identify which level-1 predictors can be modeled using level-2 variables. Second, significance testing of the variances in the random coefficients model identifies which level-1 predictors should be allowed to vary randomly across groups in the final model. The full model is called the intercept- and slopes-as-outcomes model (Raudenbush & Bryk, 2002).

The statistical model for the intercepts- and slopes-as-outcomes model with two level-1 and two level-2 predictors consists of the following equations. For the purposes of discussing the how the statistical model is written, one of the level-1 slopes will be fixed and one will vary randomly. The statistical model can be written as

**Level-1:**

\[ Y_{ij} = \beta_{0j} + \beta_{1j}(X_{1ij}) + \beta_{2j}(X_{2ij}) + r_{ij} \quad \text{where } r_{ij} \sim N(0, \sigma^2) \quad (13) \]

**Level-2:**

\[ \beta_{0j} = \gamma_{00} + \gamma_{01}(Z_{1j}) + \gamma_{02}(Z_{2j}) + u_{0j} \quad (14) \]

\[ \beta_{1j} = \gamma_{10} + \gamma_{11}(Z_{1j}) + \gamma_{12}(Z_{2j}) + u_{1j} \quad (15) \]

\[ \beta_{2j} = \gamma_{20} + \gamma_{21}(Z_{1j}) + \gamma_{22}(Z_{2j}) \quad (16) \]
where the error terms are normally distributed with means of 0 and a variance/covariance matrix. Each of the $\beta$ coefficients is written as a linear combination of the level-2 variables. The intercept, $\beta_{0j}$, and the slope, $\beta_{1j}$, are allowed to vary randomly across groups as indicated by the inclusion of error terms in the level-2 equations. The slope of the second level-1 predictor, $\beta_{2j}$, is treated as fixed. Results from statistical software will include parameter estimates for fixed effects, and variance components, and associated significance tests.

In her introduction of hierarchical linear modeling to the field of higher education, Ethington (1997) included in her analysis a special variation of the full model called the random intercepts model. This model is similar to a random coefficients model. However, only the intercept is modeled and allowed to randomly vary across groups. The estimates for each of the parameter estimates for the level-1 variables will be the same for each group and there are no level-2 predictors in the model. This is indicated by the omission of error terms for the level-2 equations that model the parameter estimates for the variables in the level-1 equation. Only the intercept is allowed to vary randomly across groups. Conceptually, the random intercept model is very similar to an OLS multiple regression, but should produce more accurate estimates of coefficients because the estimation methods and procedures are designed to accommodate multilevel data (Ethington, 1997).

**Reporting Results**

Consistent with reporting practices for single-level regression models, multiple authors recommend that results from a hierarchical linear model be reported using tables.
with accompanying narrative (Dedrick et al., 2009; Ferron et al., 2008; McCoach, 2010). These tables should include a complete list of fixed and random effect parameter estimates, standard errors, and tests of significance or confidence intervals. Dedrick et al. (2009) argued that statistical significance tests can be of limited value and difficult to interpret and seemed to argue, implicitly, in favor of confidence intervals. McCoach (2010) did not mention confidence intervals in her recommendations for hierarchical linear models. The importance of testing the model fit was addressed by multiple authors. Dedrick et al. suggested that fit indices can help determine the appropriate covariance structure and mentioned both the AIC and BIC (Akaike, 1974; Schwartz, 1978). McCoach suggested that calculating deviance provides an indicator of badness of fit and presents details on how to calculate deviance in hierarchical linear models. Ferron et al. (2008) discussed graphing slopes representing the pseudo $R^2$, which may aid in communicating information about differences across groups, but noted that this approach is not universally accepted.

**Hierarchical Linear Models as the Focus of Research**

Informed by the observation that “statistical software does not a statistician make” (Singer, 1998, p. 350), this research was grounded in a concern that the diversity of perspectives and methodologies represented by higher education research makes it difficult to critically evaluate the methodological accuracy of studies using hierarchical linear modeling. A search of higher education literature using the terms hierarchical or multilevel in the title produced few publications with titles indicating a focus on multilevel methods (Dey & Astin, 1993; Thomas & Heck, 2001). The following
describes those from the general higher education literature.

**Higher Education Literature About Multilevel Analyses**

In an essay on characteristics of “exemplary research” Smart (2005) cautions “that the use of the most recent and sophisticated analytical procedures is not necessarily the best approach” (p. 466). Smart made an unsourced observation that the popularity of LISREL in the 1980s produced articles that misapplied the technique and did not produce a meaningful reconstruction of previous findings. Making specific mention of hierarchical linear models, Smart suggests that similar events may be occurring with hierarchical linear modeling. The basis for his caution, however, was not the prior observation, but that software applications make it possible for scholars to apply hierarchical linear modeling without understanding the underlying statistical principles. The ease of use makes it possible for “under-prepared users” to conduct the analyses. Noting that advanced statistical and mathematical training is required to fully understand how to apply hierarchical linear models, Smart recommended that scholars use traditional techniques. While this may be a practical solution to a legitimate concern, the recommendation seems to contradict both the purpose of the essay, which was to identify attributes of exemplary research manuscripts and recommended statistical practices.

Pascarella (2006) offered a similar caution regarding the use of complex statistical techniques. Pascarella suggested that advanced statistical techniques, such as LISREL and HLM (here meaning the technique hierarchical linear modeling) do not provide more credible results when the quality of the data is suspect. A proponent of longitudinal pretest-posttest designs, Pascarella argued that a single campus longitudinal study should
be preferable to multi-campus, cross-sectional designs. Pascarella did, however, seem to suggest that scholars using the technique were basing claims about validity on the results of advanced statistical analyses. There appears to be no empirical evidence found in studies using hierarchical linear models to support this claim.

One of the few, perhaps only, published studies in the higher education literature to empirically compare OLS regression to hierarchical linear modeling is Astin and Denson’s (2009) study that modeled college student political identification using a set of 20 entering student characteristics and four institutional characteristics. Using a split sample, Astin and Denson first modeled political identification using a hierarchical linear model and an OLS multiple regression model on half of the sample. They then applied the resulting models to the second half of the sample to produce a predicted political affiliation and compared the predicted affiliation to the reported affiliation. While the models did differ in the accuracy of correctly classifying students as conservative or liberal, the differences in accuracy were less than 0.5%. Astin and Denson concluded that “both methods yielded essentially the same result” (p. 361). Noting a perception that journal editors were advocating the use of multilevel analysis with multilevel data, Astin and Denson then offered a series of conditions under which OLS regression is an acceptable alternative. If modeling the data temporally or to explore mediating effects using path analytic techniques, then OLS regression is “the preferred method” (p. 366). However, if any institutional level effects are significant then the data should be remodeled using multilevel techniques. They suggest this should involve using significant predictors from the OLS model to construct a hierarchical linear model with
nonrandomly varying slopes. The recommendation seems to accommodate both the strengths and weaknesses of each technique and its accompanying software. However, the results and implications, unless read closely, appeared to imply that using a more conservative p-value is an acceptable correction to the violation of the assumption of independence of observations. This is not supported by recommendations from statisticians.

Dedrick et al. (2009) conducted a study of the reporting practices of multilevel modeling from 13 peer-reviewed education and social science journals. They used a sample of 99 studies published between 1999 and 2003, and found that the majority of studies were neither experimental nor used probability sampling.

Dedrick et al. (2009) reported that the majority of studies in their research included an explicit rationale for the use of hierarchical linear models and that most acknowledged the limitations of using single-level analyses with multilevel data. The numbers of models tested ranged from 1 to 430. The majority (n = 52) examined fewer than 10 models, and it was not possible to identify the number of models tested in five studies. Only 48 studies reported modeling the fully unconditional model (described as baseline in Dedrick et al.).

Using centered variables makes it easier to interpret the coefficients and variance components. Centering of at least one level-2 variable was reported in 49 studies and at level-1 in 75 studies. The results were reported using a variety of approaches. Ninety-eight of 99 studies in the sample presented a verbal description of the fixed effects and the majority of studies (n = 92) used tables to report estimated effects. A smaller number
Presented the variance structure using a verbal description and even fewer ($n = 44$) presented the variance structure in table form.

Data considerations were the second major theme identified by Dedrick et al. (2009). Many of the data considerations discussed by Dedrick et al. were applicable to any quantitative analysis, but are important to include in reporting of results for the purpose of replication and transparency. Ten studies acknowledged or mentioned the assumptions related to normality and homoscedasticity, while only two indicated that the data were checked to verify the assumptions. It may be that study authors intended the reader to assume such testing is always included in preliminary analyses and data preparations. However, given that eliminating one of the assumptions of OLS regression (independence) is a motivating factor for multilevel analysis, it was somewhat surprising that studies did not state explicitly that the data met the conditions of hierarchical linear models. Finally, a majority of studies reported checking for missing data ($n = 80$) and applied a variety of treatments including listwise deletion, imputation, and using a proxy variable with a missingness indicator.

Dedrick et al.’s (2009) analysis of information regarding estimation and significant tests in their sample may provide evidence in support of Smart’s (2005) proposition that user-friendly software may mask a lack of understanding of the technique on the part of scholars. The majority of studies did not report key elements of estimation and significance tests. Different software applications use different numerical procedures to produce estimates. As a result, it is useful to report the software, estimation methods, and to indicate whether or not parameters are fixed or random.
Fifty-three studies did not indicate the software used. Only 15 studies stated the estimation method leaving the reader to assume that default settings in the software were applied. Perhaps reflecting an intent to interpret fixed effects as regression coefficients, 96 studies reported point estimates and significance tests. In contrast, 23 studies provided no information about the variance estimates. Standard errors were reported in 60 studies.

Dedrick et al (2009) concluded with a set of recommendations to guide reporting practices based on their findings, a book chapter by the authors, and standards published by the American Educational Research Association (AERA, 2006; Ferron et al., 2008). These recommendations are:

1. Provide a clear description of the process used to arrive at the model(s) presented. This should include a discussion of how the predictors were selected, how the covariance structure was chosen, and how many models were examined. Readers can more carefully consider the presented models if they clearly understand how the models were developed.

2. Explicitly state whether centering was used, and if used, provide details on which variables were centered and how they were centered. Knowledge of centering decisions will aid in the interpretation of regression coefficients and variance estimates.

3. Explicitly state whether distributional assumptions were considered and whether data were screened for outliers. If such checks were made, state both the method used and what was found. With this type of information, it is easier to
evaluate the credibility of the results.

4. State whether the data were complete. If they were not complete, describe the missingness and attempt to provide insight into its possible effects on the results.

5. Provide details on the analysis methods, including a statement of the software and version used, the method of estimation, whether convergence was obtained, and whether all variance and covariance estimates were admissible.

6. For any interpreted model, provide a complete list of all parameter estimates. In addition to providing critical information for interpreting the results, this helps to communicate the precise model estimated.

7. “Provide either standard errors or confidence intervals for the parameters of interest….Statistical significance tests provide limited inferential information and can be difficult to interpret when large numbers of tests have been conducted, a typical occurrence in the reviewed applications” (p. 96).

The results from the Dedrick et al. (2009) study offer insights into some of the challenges of using hierarchical linear models. First, the process for conducting an analysis consists of several steps including data preparation, understanding model specification and estimation techniques, and reading and interpreting results. Providing sufficient detail to ensure transparency and replication of results requires more pages of narrative devoted to methodology than authors may be willing to allow when hierarchical linear modeling is an analytic tool and not the focus of study. The reader, and editorial reviewer, is asked to trust that the analysis is conducted correctly and the information
presented is an accurate interpretation of results.

Despite the utility of the review and analysis, the impact of the recommendations is limited given that it was published in 2009. A citation search using googlescholar, executed in June 2012, produced only 14 publications citing this study. Only one of the 14 is in a higher education context (Cheslock & Rios-Aguilar, 2011). Second, the studies included in the sample were published between 1999 and 2003. It is possible that areas of “deficiency” identified by Dedrick et al. are now included as a matter of routine in recent publications. Finally, the sample did not include articles from any higher education journals. It was possible that a sample from higher education literature sources would produce different results.

**Higher Education Literature Using Multilevel Data**

The focus of higher education scholarship is typically on the topic of research and not the methodology, and higher education is an applied field that draws from multiple disciplines. Thus, it is not surprising that there were so few higher education publications related to the methodology of hierarchical linear models. A survey of the higher education literature that analyzed multilevel data may serve as a means to illustrate the lack of consensus regarding the topic. One of the most important questions to be asked about hierarchical linear models, and multilevel analysis in general, is under what conditions is OLS regression appropriate to use with multilevel data? Although this issue seems to have been resolved in the literature of applied statistics (Gelman & Hill, 2007), higher education research does not reflect this resolution.

There exist several examples of studies that apply OLS regression to multilevel
data. Studies dating to the early 2000s use OLS regression and rarely acknowledge the problem of independence. For example, Chang, Astin, and Kim (2004) conducted a study to model the frequency of cross-racial interaction in college students. The study used longitudinal data from the Cooperative Institutional Research Program and applied hierarchical linear multiple regression which entered variables temporally. Perhaps a motivation for the study published by Astin and Denson (2009), described earlier in this chapter, Chang et al. did find institutional characteristics significantly associated with cross-racial interactions. The multilevel structure of the data was not discussed explicitly, but was alluded to in the manner in which the general model was expressed.

As the decade progressed, a pattern seemed to emerge in the higher education research which presented a rationale for using single-level regression modeling with multilevel data. In a study of wage growth in college graduates, Thomas and Zhang (2005) used data from the Baccalaureate and Beyond to examine the effect of major and institutional quality measures on postgraduation earnings. They modeled income at two points in time. The authors commented that the data were modeled using OLS regression and hierarchical linear modeling (calling it HLM). As part of their comment they reported the results of an OLS regression and claimed that the multilevel model “yields similar results at discrete points in time” (p. 442) and that multilevel analysis “is difficult to implement when comparing differences at multiple points in time” (p. 442).

Other studies acknowledged the statistical problems of analyzing multilevel data using single-level techniques. In a study of Texas females, Riegle-Crumb (2010) sought to identify the factors that explained what she termed the “female postsecondary
advantage in matriculation among Hispanic and white youth” (p. 572). Seeming to respond directly to the commentary offered by Thomas and Heck (2001), Riegle-Crumb stated the “standard errors of all estimates are adjusted to take into account the correlations between individuals within the same cluster” (p. 580), but did not explain the procedure used to correct the standard errors.

In a study that compared survey responses from paper and online versions of the NSSE, Carini, Hayek, Kuh, Kennedy, and Ouimet (2003) included a lengthy rationale for using OLS regression instead of multilevel analysis. Carini et al. acknowledged that OLS regression will “likely produce biased standard errors” (p. 9), and cited Ethington (1997) to claim that OLS regression and multilevel analyses will produce similar results and therefore is acceptable. They also argued that the sample size for their study made significance tests “less instructive” (p. 9) and instead calculate effect sizes. It was not clear from Carini et al.’s argument that effect sizes were intended to correct for biased standard errors.

**Intraclass correlations.** Several studies in the mid- to late-2000s used the ICC as a test statistic to justify the selection of single-level or multilevel analyses. Cox, McIntosh, Terenzini, Reason, and Quaye (2010) used data from the Wabash Study and modeled two types of student-faculty interaction. They reported that they began to model their outcomes using a hierarchical linear model, but switched analytic approaches to structural equation models when the first step of the hierarchical linear model revealed that the between-group variance was only 3% of the total variance, as measured by the ICC. Cox et al. argued that the amount of between-group variance was so small that it
was practically insignificant and did not need to be considered in the later analysis.

Cole (2011) studied intellectual self-concept of African American students. Using a sample of 460 students at 96 institutions from the CIRP’s Student Information Form and College Senior Survey, Cole modeled intellectual self-concept and average college grades. Indicating a cutoff of 5% or lower to justify using OLS regression with multilevel data, the author reported an ICC of 2%. The absence of a citation to support the 5% cutoff justification made it difficult for the reader to ascertain if the conclusion was correct.

Mayhew, Seifert, and Pascarella (2012) conducted a study using data from the Wabash Study. Using a sample of 1,469 students at 19 colleges and universities, Mayhew et al. examined gains in moral reasoning development. The authors justified the use of a two-prong argument. First, they calculated and reported the ICC for their outcome measures and claimed that at 8.7% and 12.9% these were sufficiently low to justify using alternatives to multilevel analyses. The second component of their argument claimed that the focus of the study was on the experience of students. Based on the focus and ICC they concluded using a single-level analysis with the student as the unit of analysis was appropriate. Mayhew et al. acknowledged that using OLS regression resulted in underestimated standard errors and “corrected” through the use of conservative p-values. They did not provide a source for this correction.

Based on the literature reviewed for this research, the ICC has been used with some frequency to justify using OLS regression with multilevel data. Higher values for the ICC produce larger bias of standard errors (Gelman & Hill, 2007). In the motivating
example for her article, Singer (1998) calculated an ICC of .18 and commented that this means that OLS regression would produce “misleading results” (p. 330). Similar to other studies described here, this conclusion did not appear to have a theoretical or empirical basis. The absence of citations to support this conclusion makes it difficult to reconcile the practices of higher education scholars with those of applied statisticians.

The literature reviewed here motivated the development of research question three, which included a systematic analysis of arguments and corrections scholars made in single-level analyses of multilevel data. This research documented and confirmed the findings synthesized by this reviews through a systematic content analysis of a comprehensive sample of studies.

**Chapter Summary**

The literature of this review was representative of the major theoretical foundations of hierarchical linear modeling and provided a rationale for this research. The following summarizes the rationale. There is ample evidence in the statistical literature to support an assertion that multilevel techniques should be the default analytic method for multilevel data (Burstein, 1980; Hahs-Vaughn, McWayne, Bulotsky-Shearer, Wen, & Faria, 2011; Hox, 1998). Despite the statistical evidence, higher education researchers continue to use single-level regression analyses with multilevel data (Chang et al., 2004; Cox et al., 2011; Riegle-Crumb, 2010; Thomas & Zhang, 2005). Low intraclass correlations were frequently reported and used to defend the choice of single-level analyses, but did not include a citation to justify the conclusion. In the case of modeling outcomes, the argument that significant parameter estimates are the same for
both OLS regression and multilevel modeling justified using a single-level analysis. There was no clear explanation for the inconsistency between higher education practice and the recommendations of statisticians. This research included analyses that help explain these inconsistencies. More specifically, the analysis of citations related to research question one identified the most frequently used sources in studies that used hierarchical linear modeling.

There were few examples in the higher education literature that examined publication patterns and structural content of empirical studies. This review included an examination of reporting practices and recommended practices for studies employing hierarchical linear modeling. The one empirical study of reporting practices for hierarchical linear modeling in the education literature had two characteristics that justify the proposed research (Dedrick et al., 2009). However, this study did not include any higher education journals in its sample of 13 education related journals. Furthermore, the study is dated having analyzed articles published between 1999 and 2003. This research made it possible to examine reporting practices using a sample of higher education journals and include more recent literature to account for changes in practices over time.
CHAPTER THREE
RESEARCH METHODS

The questions for this research required the use of content analytic techniques to execute the analyses. Content analysis is defined as “a research technique for making replicable and valid inferences from texts…to the context of their use” (Krippendorff, 2004, p. 18). With origins in journalism and communication studies, the evolution of the method makes it applicable to research that requires the analysis of text. Contemporary methods incorporate practices from both qualitative and quantitative paradigms and consists of a specific set of characteristics to establish the credibility of research (Krippendorff, 2004).

As described in Chapter One, this research addressed three research questions related to the analysis of multilevel data by higher education scholars. The first two questions examined citation and reporting practices in studies that model continuous outcomes using hierarchical linear models. The final research question explored the rationales presented by scholars in studies that applied single-level model-based analyses on multilevel data. This research focused attention on an understudied aspect of higher education scholarship—the publications used to document that scholarship. Content analysis was the most appropriate methodology for research that uses the literature of the field as the object of analysis. No single content analysis could address the range of questions that guide this research. Instead, a series of three related designs was
developed. Each question required a unique design and plan of analysis that was consistent with content analytic techniques. Adapting Krippendorff’s (2004) description of content analysis, the study design is described for each research question according to the following format. First, the research question and hypothesis are presented. Next, the sample is defined and the coding and context units are described. With that foundation, the coding processes are described in detail along with the plan for reducing the products of coding. The section closes with a description of analyses that were performed and how results are presented. The chapter concludes with a discussion of the trustworthiness and limitations of this research.

**Content Analysis**

The historical antecedents of modern content analysis can be traced to the 19th century when scholars studied several topics through the systematic analysis of newspaper text. The earliest content analyses consisted of counting the frequency of word use and reported general trends in reporting as well as provided evidence of bias in journalism.

Content analysis as a formal research method dates to the post World War II period when scholars codified procedures (Berelson, 1952; Berelson & Lazarsfeld, 1948). The method is empirically grounded and seeks to make inferences about phenomena that cannot be accessed or observed using other methods (Krippendorff, 2004). The process of content analysis, as described by Krippendorff (2004), adopts some of the practices of scientific method, with specific, sequenced components. It is not, however, based on positivist epistemology as the process of reading text is an inherently qualitative act. The
reader mediates the text, which can bias results. However, the procedures of content analysis are designed to minimize the likelihood of this bias. The following describes the components of a content analytic study as described by Krippendorff.

**Units of Analysis and Sampling**

The first component of content analysis consists of defining the units of analysis. The process of unitizing occurs at several levels. The first unit of analysis to be defined is the *sampling* unit. The sampling unit definition is used to identify and select texts that are included in the research. The sample for this research was drawn from studies published in the top four cited U.S.-based academic journals in higher education (Budd, 1999; Budd & Magnuson, 2010). They are the *Journal of College Student Development*, *Journal of Higher Education*, *Research in Higher Education*, and *Review of Higher Education*. Studies published between January 1, 2000, and June 30, 2012, were included in the sample.

Under certain conditions, random sampling procedures can be applied to constructing a sample for analysis. Random samples may be used when all sampling units are *equally informative*, meaning that text can contribute equally to the information needed to answer the research question. The use of random sampling procedures makes it possible to address issues related to generalization of findings. Given the questions guiding this research, a sample consisting of all articles published in the four higher education journals do not meet the condition that articles are equally informative; therefore random sampling procedures could not be used. As a consequence, a series of screening procedures were applied to construct two samples that were used for this
research. The first sample consisted of articles that modeled multilevel data using hierarchical linear modeling. This sample was used in the analysis for research questions one and two and was called the HLM Studies sample. A second sample consisting of articles that modeled a continuous outcome measure from multilevel data using OLS regression was analyzed for research question three. This sample was called the Regression Studies sample.

A second analytic unit that should be defined in content analytic research is the context unit. Because one of the characteristics of content analysis is that the content is decontextualized, it is not necessary to process the text in the context of words, sentences, or paragraphs based on proximal location. The context unit defines the “size” of the text that can be attached to a code. In addition to different samples, this research used different context units for each research question ranging from a short segment of text to a full paragraph. The context unit is defined in the analytic plan for each question.

The third type of analytic unit that should be defined prior to conducting the research is the coding unit (Krippendorff, 2004) Coding units are the categories to which context units are assigned. In contrast to qualitative studies which seek out emergent themes in the narrative, content analysis can be used to identify themes in the narrative either inductively or deductively. Inductive approaches include frequency counts of words appearing in the narrative. This approach has been used in the fields of journalism and political science on newspapers and speeches. More commonly found in higher education scholarship are deductive approaches in which a collection of codes, sometimes called variables, are defined in advance of the research. Similar to practices in
variable definition in quantitative research, coding units should be defined so they are mutually exclusive and exhaustive. Every unit in the sample should be assigned a value for every coding unit in the research. The code is recorded numerically and reported using descriptive statistics. The research in this study used a mixture of inductive and deductive coding units.

**Coding Procedures**

After defining the units of analysis and the process of constructing the sample for the research, the next step of content analysis is to describe the coding procedures. The coding unit should motivate the coding procedures and incorporate both a development and execution phase. Scholars differ on how these phases are executed. Neuendorf (2001) described very specific procedures for coding that included specific steps to emphasize choices made *a priori*. This approach is appropriate for research that is based on existing theoretical frameworks. For the research presented here, on which there is no clearly identifiable theory on which to base a coding framework and, therefore, is by definition, exploratory, a more flexible approach was appropriate. Krippendorff (2004) presented an approach in which the focus is on the end result of the coding process—reflective of qualitative coding techniques. The end product, however, of both strategies included a detailed, specific process for identifying and categorizing patterns in the text. In the development phase of coding, the predefined set of coding units are applied to some texts in the sample. During this phase, the coding may be revised so it can be mutually exclusive and collectively exhaustive. The coding procedures should specify whether inductive or deductive coding strategies are used.
In addition to describing the process of coding, the coding procedures should include a description of how the products of coding are reduced to a form that can be represented as results or used in statistical analyses. Traditional qualitative research identifies themes and presents results using the language of the text to illustrate or convince the reader of the credibility of the theme (Glesne, 1999). Content analysis, in contrast, reduces the products of coding to numerical form. Codes, for example, may represent the presence or absence of a particular characteristic or phenomena of interest. This code is assigned to every text in the sample. In this research, for example, every article is assigned a code that represents the source journal and another that represents the year of publication. Other codes may represent a quality of a characteristic that is present in the text and are assigned each time this characteristic is found in the text, which may be zero if it is not present, one if it occurs one time, or greater than one if it occurs multiple times in the text. The coding procedures should include a description of how the coding is transformed and analyzed into the form that is presented as the results of the research. The level of detail incorporated into a description of the coding procedures in content analytic studies is intended to increase the rigor and credibility of design. One of the distinctive characteristics of this research is that coding and data reduction procedures vary by research question.

**Analyses and Reporting**

Research is not complete unless the results are documented. A complete analytic plan should include a description of all relevant analyses and a brief description of how the results of these analyses are presented. Collectively, the components of content
analysis describe a design that is systematic, replicable, and credible. These components will guide the description of the proposed research.

**Study Design**

As stated in Chapter One, the purpose of this research was to examine how higher education scholars have applied and communicated information regarding methodology in studies that analyzed multilevel data. In 2005, Smart postulated that higher education scholarship may not be applying advanced statistical techniques appropriately. This research examined the reporting practices of higher education scholars who are analyzing multilevel data to answer questions about phenomena in the higher education context. One of the key contributions of this research was that it provides a baseline description of how this technique has been applied in higher education research as only one empirical study focused exclusively on higher education literature was identified by a search of the literature. Research on the nature of academic disciplines and discourse communities indicated the literature of the field reflects the values, beliefs, and communal knowledge of a topic (Dressel & Marcus, 1982; Lattuca, 2001). Given the general absence of literature that explicitly addresses methodological issues regarding hierarchical linear models, one way to explore higher education’s understanding of this analytic method was to examine how information about the technique was communicated in the studies that used it.

**HLM Studies Citation Analysis**

The purpose of research question one was to understand the body of literature cited by scholars when writing about hierarchical linear models. As stated previously,
the literature about multilevel analysis, and hierarchical linear models specifically, is represented in a broad cross-section of disciplinary and field perspectives, including epidemiology, geography, political science, and education. It seemed reasonable to assume that, in a field such as higher education, which draws from a similarly broad set of disciplinary perspectives, one would also find the varied perspectives of hierarchical linear modeling represented. One approach to examining the perspective represented in higher education scholarship was to examine the sources cited by scholars in written reports of the research.

**Research question one.** What methodological sources have been cited by higher education scholars who have published studies that used hierarchical linear models?

**Hypothesis one:** *Citation analysis will show that scholars draw from a limited set of resources.* In addition, the sources are not representative of the variety of disciplinary perspectives using hierarchical linear models. Prior citation analyses have included all the citations in a published article and have not differentiated among types of citations (Budd & Magnuson, 2010; Snyder, Cronin, & Davenport, 1995; Tight, 2008). Citation analyses of general higher education literature suggested that scholars relied on the major higher education journals and tended to use the same sources repeatedly in their writing (Budd & Magnuson, 2010). While some sources may be sufficiently broad in scope or described a theory used frequently to study a topic (e.g., Astin, 1993; Pascarella & Terenzini, 1991, 2005), these results may also indicate that scholars were not exploring literature outside the field. There exists a large body of methodological technical literature related to multilevel analysis published in sources other than higher education
journals. As a result, it was not known if methodology related citations are similarly limited. The analysis for research question one tested the hypothesis that scholars are citing a relatively narrow set of sources when reporting the results of studies using hierarchical linear models. Confirmation of this hypothesis may provide additional support for Smart’s (2005) concern that advanced statistical techniques are misapplied in higher education research.

**Analytic plan.** Research question one applied citation analysis to describe the sources higher education scholars cited in empirical studies that use hierarchical linear models.

**Sample description.** As indicated in the description of content analysis, two samples were constructed for this research. All samples were drawn from studies published in four journals: *Journal of College Student Development, Journal of Higher Education, Research in Higher Education, and Review of Higher Education*. Studies published between January 1, 2000, and June 30, 2012, were considered for inclusion in the sample(s). To identify those studies that used hierarchical linear models, a series of electronic and manual reduction procedures were employed. The first reduction was conducted using search functions of journal databases. Each article published during the specified time frame was searched electronically using the search terms *multilevel OR multi-level OR hierarchical OR HLM* using the full-text search command.

A search of the *Review of Higher Education* yielded 88 articles. *Research in Higher Education* and *Journal of Higher Education* had 168 and 96 eligible articles, respectively. A similar search of the *Journal of College Student Development* yielded
105 possible studies for the years 2003 through June 2012. The *Journal of College Student Development* online database did not allow for full text searches for years before 2003. All articles published between 2000 and 2002 were reviewed manually. The manual review of articles from the *Journal of College Student Development* yielded one article to include in the sample for research questions one and two and four articles for the sample for research question three. The key word search identified a total of 467 articles that included at least one of the key words.

The next stage of the sample screening process incorporated both automated and manual screening. An article was included in the final sample for research question one if it reported the results of a study that used hierarchical linear modeling. Because the term *hierarchical linear modeling* has been used as a generic term to represent multilevel models, the outcomes measures were also reviewed for studies meeting the first condition to ensure the article is using the statistical definition of a hierarchical linear model.

Each of the articles in the reduced list \((n = 467)\) was loaded into Atlas.ti software. Using a feature of the software that allows one to use automated coding to assign codes to text strings, a series of automated coding procedures was performed to make the second article reduction more efficient. The following codes were assigned for the data reduction process: *hierarchical, HLM, regression, multilevel, multi-level, review, and qualitative*. Each of the 467 articles was scanned with particular attention paid to the text segments assigned to using automated coding and the abstract, if included. Final determination of article status was made based on both exclusion and inclusion criteria. Reasons for exclusion included, but were not limited to, the following: literature and
book reviews, case and/or qualitative studies, other quantitative methods, single-campus studies, single-level regression models, and other multilevel analyses. The reason for exclusion or confirmation of inclusion was recorded in an Excel spreadsheet. The final sample consisted of 60 articles across the four journals. A listing of references for the sample is included in Appendix A.

Coding procedures. The coding for research question one applies citation analysis, which may be considered a variant of content analysis (Eom, 2009). The context unit is defined as a unit of text that is identified for analysis. The first context unit is the citation included in the text itself. The citation represents a source material to which the author of the publication makes attribution for a concept, fact, or quotation. In the coding process for this study, the citation was used for identification in the context of the narrative. The coding unit, defined as the categories to which context units are assigned, for this study was the identification of a citation as related to hierarchical linear modeling.

The coding process for this research consisted of the following procedures. An Excel spreadsheet was created and each article was coded with the journal source, publication year, and total number of sources in the reference list. Next, all citations appearing under a heading related to the methodology of a study were identified and the entire text was reviewed to identify statements related to methodology, specifically multilevel data or multilevel analysis. Next, the reference list for a study was reviewed to identify those sources that were known to be related to hierarchical linear modeling or statistics. For example a study published in the Journal of Educational and Behavioral
Statistics would qualify for inclusion in the citation analysis. Similarly, the article, *Intraclass Correlation Values for Planning Group-Randomized Trials in Education* (Hedges & Hedberg, 2007) would be included in the analysis for research question one. This produced a list of methodological sources for each article. The total number of sources for each article was recorded in an Excel spreadsheet. A matrix that mapped each source to all articles citing it was also created using an Excel spreadsheet. Finally, each formatted source in the list was assigned codes that represented the type of source: book or book chapter, journal article, technical manual, internet source, or other.

**Analysis and reporting.** The analysis for research question one consisted of primarily descriptive statistics. Analyses included descriptions of source allocation by journal. Next the sources used were examined in more detail. The methodological sources cited most frequently in the sample articles were reported. Finally, the type of sources were analyzed and reported.

As the coding and analysis was conducted for this research question, stylistic differences in how authors structured methods sections of articles made it necessary to amend the coding and reporting plan for this research question. These amendments, which produced results more accurately reflecting the intent of the research question are described in Chapter Four. The final codebook used in the research for research question one is included in Appendix B.

**HLM Studies Content Analysis**

The purpose of research question two was to describe the content scholars included in published studies that report the results of studies that used hierarchical linear
modeling. As described in Chapter Two, only recently has there been documented guidance on how to report results of empirical research using hierarchical linear models. The analysis for research question two extended both the work completed by Dedrick et al. (2009) which examined the methodological content reported in studies using hierarchical linear modeling in general education and psychology journals published between 1999 and 2003 and that of Cheslock and Rios-Aguilar (2008, 2011) that examined studies published in the *Journal of Higher Education* and the *Review of Higher Education*. This research also provides a more current perspective on reporting practices because it included published studies between 2000 and 2012 and expands that understanding to include *Research in Higher Education* and the *Journal of College Student Development*. The inclusion of these additional journals is relevant because *Research in Higher Education* had the greatest proportion of studies to contribute to the present study (33 of 60 total).

**Research question two.** Using Dedrick et al.’s (2009) analytic framework for examining the narrative content of studies using hierarchical linear modeling, what methodological issues are included/omitted in narratives of studies using the technique?

**Hypothesis two.** The narrative related to methodology in studies using hierarchical linear models provides insufficient information to the reader to determine if the application and interpretation of results are accurate. In their research on studies using multilevel analyses in a selection of general education journals, Dedrick et al. (2009) identified several inconsistencies in the reporting practices applied to studies that use hierarchical linear models. Cheslock and Rios-Aguilar (2008) reported similar
findings in a study of 14 multilevel analyses published in the *Journal of Higher Education* and the *Review of Higher Education*. Several of the studies reviewed in Chapter Two included justifications for the use of single-level analyses with multilevel data that were not supported by the technical and methodological literature (e.g., Gelman & Hill, 2007). In combination, this evidence supports a hypothesis that a comprehensive analysis of hierarchical linear modeling studies from higher education journals will identify similar deficiencies in reporting practices.

**Analytic plan.** Research question two used the *HLM Studies* sample. This sample consists of 60 articles, representing all studies that reported results of hierarchical linear modeling published between January 1, 2000 and June 30, 2012, in *Journal of Higher Education, Journal of College Student Development, Research in Higher Education*, and *Review of Higher Education*.

**Coding procedures.** The study design for research question two incorporated practices of traditional content analysis. Prior to the conducting this research the *context unit*, defined as the length of the text string assigned to codes, was determined to be a sentence. This was later amended to be an entire paragraph because it was sometimes necessary to interpret a sentence in the context of those adjacent to it in order to make the appropriate determination with respect to coding.

The *coding units* used in research question two were adapted from a checklist developed by Dedrick et al. (2009). As described in Chapter Two, Dedrick et al. developed a coding framework based on issues identified from a “technological and methodological review of the multilevel literature” (p. 85). This framework included
four areas: model development and specification, data considerations, estimation, hypothesis testing and statistical inference. In preparation for this research the checklist used by Dedrick et al. (n.d.) was obtained from one of the authors. This checklist included a total of 73 discrete codes. Because the focus of this research is on what has or has not been reported in publications, the Dedrick et al. ’s checklist was adapted and reduced to 40 codes. Table 2 lists the variables coded in this research by the themes identified by Dedrick et al. and described in the review of the literature. The complete codebook for this study can be found in Appendix C.

Table 2. Guiding Questions Used to Create HLM Reporting Codebook

<table>
<thead>
<tr>
<th>General Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What best describes the study type? Is a rationale (and/or advantage(s)) provided for using multilevel methods in the study?</td>
</tr>
<tr>
<td>2. Thoroughly describe the data set, including scope if known.</td>
</tr>
<tr>
<td>3. What type of sampling was used?</td>
</tr>
<tr>
<td>4. How many level-1 units per level-2 unit?</td>
</tr>
<tr>
<td>5. How many level-2 units?</td>
</tr>
<tr>
<td>6. How well was the distribution of level-1 units across level-2 units addressed?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How many models are examined in the study?</td>
</tr>
<tr>
<td>2. How well was the number of models in this study communicated?</td>
</tr>
<tr>
<td>3. Were baseline models run?</td>
</tr>
<tr>
<td>4. How were the predictors selected?</td>
</tr>
<tr>
<td>5. Were there more than one set of predictors for each dependent variable?</td>
</tr>
<tr>
<td>6. Were interactions examined in the presented models?</td>
</tr>
<tr>
<td>7. How was the covariance structure of the model(s) specified?</td>
</tr>
<tr>
<td>8. Was there centering of variables at level-1?</td>
</tr>
<tr>
<td>9. Was there centering of variables at level-2?</td>
</tr>
<tr>
<td>10. How were the fixed effects in the model communicated?</td>
</tr>
<tr>
<td>11. How were the variance structures in the model communicated?</td>
</tr>
<tr>
<td>12. Which methods/approaches were used to evaluate generalizability?</td>
</tr>
</tbody>
</table>
Each article in the sample was read and coded on electronic copy of the article using Atlas.ti 7.0 software designed for qualitative analysis. Codes were also recorded manually on a paper checklist (Appendix C). The article identified via the manual review of articles in the Journal of College Student Development was annotated on a paper copy of the article and coding recorded on a paper checklist. Coding recorded on the paper checklists were transferred to an Excel spreadsheet for further analysis.
**Analysis and reporting.** Most codes in the codebook for this research question represented the presence or absence of a specific element in the narrative of an article in the analytic sample. Once the code was identified in the text, additional occurrences were not recorded. Results are presented in a series of tables grouped according to the following: (a) study characteristics, (b) data preparation, (c) model specification, and (d) estimation and testing.

**Regression Studies Content Analysis**

As discussed in the review of the literature, prior to the mid-1990s scholars wishing to apply model-based approaches to multilevel data were “forced” to use single-level modeling techniques such as OLS regression. This was due to limitations in both statistical theory and tools to perform the complex calculations. These limits were generally acknowledged in the statistical literature, but rarely addressed in applied scholarship. With the introduction of software such as HLM and mlWIN, the technological limitations were reduced. It may seem reasonable to assume scholars would pursue one of two analytic paths. The first path would be to adopt multilevel approaches to analyzing nested data and models with a multilevel structure. The second path would be to continue to use single-level modeling techniques, but acknowledge the issues this approach presents in terms of the validity/robustness/credibility of results and provide evidence to support the selection of the approach. Research question three explored how scholars who have adopted the second approach have addressed the issue.

**Research question three.** What reasons do scholars give in published articles to justify the use of single-level modeling approaches on complex data?
Hypothesis three. *Thematic analysis of studies that use single-level models on multilevel data will show a lack of consensus regarding accepted rationales for the persistent use of single-level analyses with multilevel data.* Similar to the background for research question one, a search for relevant literature produced no studies in which the object of analysis is the justification or rationale for methodological choices made by authors in the process of conducting their scholarship. Instead rationales are presented in the context of empirical studies. As such the design for research question three was exploratory. It was hypothesized that the analysis of articles in the sample for this research would show that scholars would present a variety of rationales.

Analytic plan. The *HLM Studies* sample analyzed for research questions one and two was not appropriate for the focus of research question three. Using the inclusion and exclusion criteria recorded in the Excel spreadsheet when constructing the *HLM Studies* sample, the 407 articles excluded from the first sample were scanned a second time to identify studies that modeled a continuous outcome variable using a single-level regression model on multilevel data. Conceptually, the inclusion criteria identified studies with data structures and outcome measures that could be used with hierarchical linear modeling. Using the results of Maas and Hox’s (2005) study that found multilevel data with fewer than 10 level-2 groups can be managed with statistical accuracy using single-level regression techniques, studies based on fewer than 10 groups were excluded from the sample. The final sample for research question three consisted of 50 articles across the four journals, including four articles identified from the manual screen of
articles published in the *Journal of College Student Development* between 2000 and 2002. A listing of all articles for this sample is included in Appendix D.

**Coding procedures.** The context unit used in the analytic plan for research question three was a paragraph. This choice was based on two factors. First, of the three analytic plans described here, the research question made it necessary to interpret the language of the authors in context. As such, using an entire paragraph made it possible to infer meaning using only the coding unit. Second, during the process of coding, the context unit for research question two was expanded from a sentence to the paragraph containing the sentence. It seemed reasonable to use the larger context unit for the analysis of research question three. The coding units for research question three include the following: explicit acknowledgement that the data are multilevel in the narrative, whether or not justification for a single-level analysis is presented, and if so, a description of the various justifications. The codebook used in the analysis for research question three is included in Appendix E.

All available electronic versions for the articles in the sample were loaded into Atlas.ti and coding was completed and recorded using the software. Each article was reviewed and codes assigned to paragraphs containing the information relevant to the code. Unlike the majority of coding in this research, which was conducted using deductive approaches, the procedures for identifying reasons scholars used to justify the use of a single-level technique was conducted using inductive practices. Again, reporting tools available in Atlas.ti were used to create a mapping of codes and the articles to which they have been assigned. This map was entered into an Excel spread sheet for
analysis. The four articles identified via the manual review of the *Journal of College Student Development* were coded on paper versions of each article. The products of that coding were entered in the Excel spreadsheet.

**Analysis and reporting.** The coding for research question three represented two types of coding strategies: the presence or absence of a characteristic (e.g., acknowledgment that data are multilevel in structure) and the identification of themes related to the code (e.g., arguments presented by the authors to justify the use of a single-level analysis on multilevel data). Because the focus of this research had been the number of articles that have the code assigned at least one time codes were transformed into dichotomous variables to indicate the presence or absence of the code in the article. Results are presented as a series of tables representing the incidence of occurrence across all articles in the sample. Coding representing the justifications were presented both in table and narrative form.

**Trustworthiness of the Data and Analyses**

This research applied content analyses to a manually constructed sample of studies that use hierarchical linear modeling. As evidenced by the study design, content analysis incorporates characteristics and procedures from both quantitative and qualitative traditions (Krippendorff, 2004). Reliability of content analyses can be evaluated using a number of approaches which are incorporated into the design of the study (Krippendorff, 2004).

Stability of coding is the extent to which coding procedures produce the same results (Krippendorff, 2004). For each of the research questions, refinements to the
codebook and coding processes were made if it became evident that the coding units, or variables in the codebook were not capturing the precise content to accurately answer each research question. Each article included in the research coding was verified during the process of reducing the coding products to numerical values used in the analyses. All coding was confirmed against the relevant text when codes were transferred from paper or electronic codes to Excel spreadsheets.

Efforts were made to make the coding and analyses rely on manifest content which minimizes the likelihood of error due to misinterpretation. The focus on manifest content can be found in the design of research questions one and two. The analysis for research question three made it necessary to identify latent content, meaning that the coding process requires a greater level of interpretation of text. Coding categories were developed in accordance with recommended guidelines for content analysis (Krippendorff, 2004). Categories were exhaustive and mutually exclusive. These guidelines were also used to develop inclusion and exclusion criteria for reviewing eligible articles.

**Ethical Issues Related to this Research**

Finally, according to university policy, a research project “which will involve human subjects…must be submitted to the IRB for review” to deem whether it will be one of minimal risk to participants as determined by the U.S. Federal Government Department of Health and Human Services (2009) regulation 45 CFR § 46.10. This regulation states the probability and magnitude of harm or discomfort anticipated in the research should not be greater in and of themselves than any ordinarily encountered in
daily life, or during the performance of routine physical or psychological examinations or tests. This research did not collect data from human subjects, nor did it include secondary analysis of data collected from human subjects. Therefore, IRB approval was not required to conduct the study.

Chapter Summary

This chapter provided a detailed description of the proposed design for this research. A description and rationale of the selected methodology, content analysis, was presented. The sample for each research question was described. Krippendorff’s (2004) framework for content analysis, the coding, data reduction, analysis and reporting plan were described for each research question. Chapters Four and Five present the results of the research and discussion of the findings.
CHAPTER FOUR

RESULTS

The purpose of this research was to understand how scholars communicate the results of research that model multilevel data using a hierarchical linear model. There are few examples of published research that examine the reporting practices in the context of scholarly publication. As a consequence, the design for this research intentionally explored different aspects of scholarly publication related to hierarchical linear modeling. The research questions required a study design using two different samples of articles and three variations of content analytic approaches. Content analysis is a research technique that seeks to minimize the interpretations made when analyzing textual data. This was achieved through a process of specifying and defining variables and procedures a priori. This was intended to reduce the effect of text being mediated through the knowledge and experience of the person performing the analysis.

The first research question, which asked what sources scholars are citing in their publications, was answered using a variant of citation analysis. The second research question, which sought to define the content scholars include when writing-up the results of hierarchical linear modeling research was explored using a content analytic framework developed by Dedrick et al. (2009). The final research question sought to understand why scholars continue to use single-level analytic techniques to analyze multilevel data. This question was answered using a sample of scholarly articles.
published in higher education journals and a content analysis of the arguments made, if any, for the selection of a single-level technique. This chapter presents a brief description of the process of creating the samples and description of the characteristics of the samples followed by detailed results for each of the analyses performed.

Sample Characteristics

Articles published between January 1, 2000, and June 30, 2012, in the *Journal of College Student Development, Journal of Higher Education, Research in Higher Education*, and the *Review of Higher Education* served as the source of data for this research. Two samples of articles were needed for this study. The first sample included scholarly articles that reported the results of studies using the technique “hierarchical linear modeling” and was described as the *HLM Studies* sample. The second sample included articles that applied a single-level variant of a regression model on multilevel data and was described as the *Regression Studies* sample. Because the focus of the research questions was on articles that met certain criteria, it was neither possible to predict how many articles would meet that criteria nor to use random sampling strategies to create the samples. As a result, a series of automated and manual screening procedures were developed and applied that identified all articles meeting the inclusion criteria from the specified journals during that time period. A detailed description of the screening procedures can be found in Appendix F. Sample one, *HLM Studies*, consisted of 60 articles. Sample two, *Regression Studies*, consisted of 50 articles. The characteristics of each sample are described in the next section.
Sample One: HLM Studies

A total of 60 articles were included in the analysis for research questions one and two. A majority of these articles were published in the journal *Research in Higher Education* \((n = 33)\). Five articles were published in *Journal of College Student Development*, which contributed the smallest number of articles to the sample. Table 3 summarizes the distribution of publications by journal source. Looking at the distribution of articles by publication year, it appeared that the number of published articles that use hierarchical linear models has increased over the publication time frame (2000 to June 2012), but the upward trend has not been consistent. The number of articles published each year ranged from a low of two (2001, 2003, 2004, 2005, 2009) to a high of 11 published in the first half of 2012. Figure 1 summarizes the distribution of articles in this sample by publication year. A total of 86 scholars are represented by the articles in the *HLM Studies* sample, with the majority \((n = 64)\) authoring one article. The most prolific scholars related to hierarchical linear modeling studies in higher education journals were Paul Umbach, George Kuh, and Stephen Porter with 10, eight, and five articles respectively.

Table 3. Distribution of HLM Studies Sample Articles by Journal Source

<table>
<thead>
<tr>
<th>Journal</th>
<th>(n^a)</th>
<th>(%^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Research in Higher Education</em></td>
<td>33</td>
<td>55</td>
</tr>
<tr>
<td><em>Journal of Higher Education</em></td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td><em>Review of Higher Education</em></td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td><em>Journal of College Student Development</em></td>
<td>5</td>
<td>8</td>
</tr>
</tbody>
</table>

\(^a\)\(N = 60\). \(^b\)Total may not equal due to rounding.
Sample Two: Regression Studies

The screening procedures described in a prior section and detailed in Appendix F produced a total of 50 articles meeting the inclusion criteria for research question three. The Regression Studies sample consists of articles that used a single-level regression model to model a continuous outcome measure using multilevel data. A majority of these articles were published in the journal Research in Higher Education ($n = 17$). Seven articles were published in Journal of Higher Education, which contributed the smallest number of articles to the sample. Table 4 summarizes the distribution of publications by journal source. The distribution of regression articles by publication year did not provide the same level of information that it did for the HLM Studies sample other than to suggest scholars did, in fact, persist in their use of the technique. The number of articles published each year ranged from a low of one in 2001 to a high of 10 in 2003. Figure 2 summarizes the distribution of articles in this sample by publication year.
Ninety-two authors are represented in the articles included in the *Regression Studies* sample. Seventy-one scholars authored a single paper in the sample during the time frame. Ernest Pascarella has authored the most articles included in the *Regression Studies* sample with seven publications between 2000 and 2012. Mayhew and Seifert were the next most published authors in the *Regression Studies* sample with four publications each included in the sample.

Table 4. Distribution of Regression Studies Sample Articles by Journal Source

<table>
<thead>
<tr>
<th>Journal</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Research in Higher Education</em></td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td><em>Review of Higher Education</em></td>
<td>15</td>
<td>30</td>
</tr>
<tr>
<td><em>Journal of College Student Development</em></td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td><em>Journal of Higher Education</em></td>
<td>7</td>
<td>14</td>
</tr>
</tbody>
</table>

\(^a N = 50. \) \(^b \)Total may not equal 100 due to rounding.

Figure 2. Distribution of Regression Studies Sample Articles by Publication Year

*Note: 2012 includes only articles published before June 30, 2012*
Results

Because the design of this research consisted of three distinct analyses on two different samples, the results are organized by research question. The results of each analysis are presented and interpreted in the context of the hypothesis. A synthesized interpretation of results is presented in the discussion of Chapter Five. Each section begins with a summary of the research question, hypothesis, and methodology. Because this research included a qualitative component and was exploratory, refinements to the analytic method are described. Next, the results for each analysis are presented. Each results section concludes with a summary of results and determination of the hypothesis.

Research Question One: HLM Studies Citation Analysis

Hierarchical linear modeling is a somewhat recent development in statistical techniques dating to the 1980s, and one that has application to a large number of disciplines and fields (e.g., education, geography, political science, demography, epidemiology). As a consequence, the literature on the technique has been dispersed across these multiple fields and our understanding of the limitations of software and how it performs in different methodological contexts is similarly dispersed in the literature of multiple fields and disciplines. Because little is known about the citation practices of scholars related to research methodology, the first question in this study sought to identify the sources higher education scholars use when reporting the results of studies that use hierarchical linear modeling. Citation analysis, a technique sometimes used to identify those sources with the greatest impact on a field was adapted for this study to identify the sources used by scholars in publications that reported the results of research that used hierarchical linear modeling. Using citations as the starting point, a data set
consisting of sources related to the methodology, hierarchical linear modeling, was constructed using the articles in the HLM Studies sample. It was hypothesized that the results would yield a small collection of resources that are cited with high frequency. If true, this could represent a collective understanding of hierarchical linear modeling with a relatively high level of coherence.

**Refinements to method.** The analytic plan for this research consisted of identifying and coding citations in the text that related to methodology. The HLM Studies sample was used for this analysis. Each article was reviewed to identify citations related to methodology that were to be included in the analysis. The proposed analytic plan stated specifically that the analysis would include citations contained in the text under the headings related to methodology (e.g., Methods, Plan for Analysis). It became evident early in the analytic process that stylistic differences in narrative styles of authors would result in a source list that included citations not directly related to hierarchical linear modeling. For example, some scholars included citations related to the selection of variables included in the model in narrative under the subheading Methods. Other authors included that content under a subheading describing the review of relevant literature. To clarify and eliminate sources extraneous to the research question, a set of codes to represent how the citation was used in the text was developed. Of this code list, four were determined to be directly related to hierarchical linear modeling. They were: (a) analysis, (b) ICC, (c) justification, and (d) unit of analysis. Several additional codes were identified through the analyses but were not reported here. These included citations associated with data preparation, the instrument used to collect data, variable
specification, and content related to self-reported data. The focus of the research question was the basis for excluding these citations. These codes were not specific to methods and analyses of hierarchical linear models. This refinement eliminated over 100 sources and may have produced results that were not reflective of the intent of this research.

The analysis also revealed multiple citations with errors. These errors typically misstated the list of authors or publication year (e.g., Hu & Kuh, 2003; Lietz & Matthews, 2010). For example, Raudenbush and Bryk’s 2002 text, *Hierarchical Linear Modeling*, was cited with an incorrect publication year in at least two articles in the sample (Lietz & Matthews, 2010; Seifert & Umbach, 2008). Sources with errors were corrected prior to conducting the final analysis.

**Results for research question one.** The analysis for research question one identified a total of 113 different sources across the 60 articles included in this research. The complete list of cited sources is reported in Appendix G. Of these 113 sources, 31 were cited in more than one article in the sample. Eighty-two sources were cited by exactly one article in the sample. While this identified the sources used, it was not sufficient information to test the hypothesis that scholars draw from a limited set of resources when writing about hierarchical linear modeling. Further analyses examined the frequency with which sources were cited, the type of sources used, and the authors’ intended purpose for the citation.

**Frequencies of cited sources.** The first analysis conducted on the data set was to count the number of articles citing the source. The most frequently cited source was
Raudenbush and Bryk’s 2002 text on hierarchical linear models, which was cited in 34 articles in the sample. The 2002 version was the 2nd edition of this book as the first edition was also a frequently cited source by scholars. Bryk and Raudenbush’s 1992 original edition of *Hierarchical Linear Models* was cited in 17 articles and was the third most frequently cited source in the data set. Ethington’s 1997 chapter, the first publication to specifically discuss multilevel models in a higher education research context, was cited by 20 articles in the sample and was the second most frequently cited source. Combined, the 1992 and 2002 texts written by Bryk and Raudenbush were cited in 49 of the 60 articles in the *HLM Studies* sample. Table 5 lists the 18 sources cited in at least three articles in the *HLM Studies* sample. It was interesting to observe that, after the top three sources, the rate of citation dropped significantly. Heck and Thomas’s (2000) book on multilevel modeling, *An Introduction to Multilevel Modeling*, and Kreft and de Leeuw’s 1998 text, *Introducing Multilevel Modeling*, are each cited in nine articles and Luke’s (2004) text on multilevel modeling and Thomas and Heck’s (2001) article on the analysis of complex survey data, published in *Research in Higher Education*, were cited in seven articles. These results could be interpreted to mean that scholars using hierarchical linear models have reached an implicit agreement on what are the appropriate sources related to the use of the technique. However, only seven sources are cited by greater than 10% of articles in the *HLM Studies* sample and there are 80 sources cited in only one article. This suggested the citation process was more complex and additional analyses were warranted. A table of all sources by frequency of citation
occurrence is reported in Appendix H. The next section explores the characteristics of the cited sources.

Table 5. Sources Cited by at Least Three Studies in the HLM Studies Sample

<table>
<thead>
<tr>
<th>Source</th>
<th>$n^a$</th>
<th>Source Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raudenbush &amp; Bryk (2002)</td>
<td>34</td>
<td>Book</td>
</tr>
<tr>
<td>Ethington (1997)</td>
<td>20</td>
<td>Chapter</td>
</tr>
<tr>
<td>Bryk &amp; Raudenbush (1992)</td>
<td>17</td>
<td>Book</td>
</tr>
<tr>
<td>Heck &amp; Thomas (2000)</td>
<td>9</td>
<td>Book</td>
</tr>
<tr>
<td>Kreft &amp; de Leeuw (1998)</td>
<td>9</td>
<td>Book</td>
</tr>
<tr>
<td>Thomas &amp; Heck (2001)</td>
<td>7</td>
<td>Article</td>
</tr>
<tr>
<td>Burstein (1980)</td>
<td>5</td>
<td>Article</td>
</tr>
<tr>
<td>Hox (2002)</td>
<td>4</td>
<td>Book</td>
</tr>
<tr>
<td>Porter &amp; Umbach (2001)</td>
<td>4</td>
<td>Article</td>
</tr>
<tr>
<td>Porter (2005)</td>
<td>4</td>
<td>Chapter</td>
</tr>
<tr>
<td>Singer (1998)</td>
<td>4</td>
<td>Article</td>
</tr>
<tr>
<td>Snijders &amp; Bosker (1999)</td>
<td>4</td>
<td>Book</td>
</tr>
<tr>
<td>King (1999)</td>
<td>3</td>
<td>Book</td>
</tr>
<tr>
<td>Pascarella &amp; Terenzini (1991)</td>
<td>3</td>
<td>Book</td>
</tr>
<tr>
<td>Robinson (1950)</td>
<td>3</td>
<td>Article</td>
</tr>
</tbody>
</table>

$^a$The complete list of sources is listed in Appendix G.

Source characteristics. The two texts by Bryk and Raudenbush (1992, 2002) were the primary source scholars cited when reporting the results of research that used hierarchical linear models. It was less obvious from a simple frequency count, that the most frequently cited sources were books on the topic (Bryk & Raudenbush, 1992; Heck & Thomas, 2000; Hox, 2002; Kreft & de Leeuw, 1998; Luke, 2004; Raudenbush & Bryk, 2002). This result combined with the large numbers of sources that were cited only one time in the sample articles suggested additional analysis of the sources was warranted.

The first analysis conducted on the source characteristics was to classify sources by type. The following list of source types were used for this analysis: book, book chapter,
conference paper, journal article, report, technical manual, and webpage/other. Of the 113 cited sources, journal articles were the greatest percentage of sources cited at 46.0% \((n = 52)\). Webpages/other were the least represented on the source list with two sources (1.8%). Books comprised 23.9% \(n = 27\) of all cited sources. Finally, book chapters represented 9.7% \(n = 11\) of the cited sources. Table 6 summarizes the distribution of source types identified by this analysis.

Table 6. Types of Sources Cited in HLM Studies Sample Articles

<table>
<thead>
<tr>
<th>Source Types</th>
<th>n(^a)</th>
<th>(%)(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal Article</td>
<td>52</td>
<td>46</td>
</tr>
<tr>
<td>Book</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td>Technical Manual</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Book Chapter</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Conference Paper</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Report</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Webpage/Other</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

\(^a\text{N} = 113.\quad ^b\text{Totals may not equal 100 due to rounding.}\)

The high incidence of journal articles as a percentage of all sources led to two additional analyses. First, the sources of type “journal article” were analyzed to create a list of journals from which the sources were selected. The 52 articles in the source list were published in 33 different journals. *American Educational Research Journal, Educational Evaluation and Policy Analysis, Journal of Educational and Behavioral Statistics,* and *Research in Higher Education* contributed the greatest number of articles at three articles per journal. The contributing journals were also coded for their discipline/field. Higher education was the focus of five journals on this list. In addition to the four journals used to build the samples for this study, *New Directions for Institutional Research,* was also a source for citations. Eleven general or other education
journals were represented on the list. Five journals identified by this analysis included the terms “methods,” “methodology,” or “statistics.” Public health, management, economics, and medicine are discipline-specific journals represented in the list. Table 7 summarizes the source journals, the number of sources from that journal in the data set, and the discipline of the journal. Scholars did not appear to draw their sources from a narrow set of journal sources, which could be interpreted to mean that they have examined a broad literature base related to hierarchical linear models. However, the disciplinary base for the journal sources was primarily general or higher education, which suggests scholars were searching for literature that is primarily in their field. Given that hierarchical linear models have evolved across multiple disciplines and that statistical and methodological literature remain a minority source for citations, it may be that higher education scholars were not searching the general statistical literature as a knowledge source for hierarchical linear models.

Table 7. Journal Source Frequency of Articles Cited in HLM Studies Sample

<table>
<thead>
<tr>
<th>Journal Title</th>
<th>Discipline</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Educational Research Journal</td>
<td>Education</td>
<td>3</td>
</tr>
<tr>
<td>American Psychologist</td>
<td>Psychology</td>
<td>1</td>
</tr>
<tr>
<td>American Sociological Review</td>
<td>Sociology</td>
<td>1</td>
</tr>
<tr>
<td>American Statistician</td>
<td>Statistics</td>
<td>1</td>
</tr>
<tr>
<td>Asia Pacific Education Review</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td>Australian Journal of Education</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td>British Journal of Educational Psychology</td>
<td>Ed Psych</td>
<td>1</td>
</tr>
<tr>
<td>Contemporary Educational Psychology</td>
<td>Ed Psych</td>
<td>1</td>
</tr>
<tr>
<td>Current Issues in Education</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td>Educational Evaluation and Policy Analysis</td>
<td>Education</td>
<td>3</td>
</tr>
<tr>
<td>Educational Research and Evaluations</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td>International Journal of Educational Research</td>
<td>Education</td>
<td>1</td>
</tr>
<tr>
<td>Journal of College Student Development</td>
<td>Higher Ed</td>
<td>2</td>
</tr>
</tbody>
</table>
The purpose of citation. The final analysis conducted for research question one examined the coding used to describe the authors’ intent when including methodology-related citations in their research. A citation is embedded in a sentence and supports directly, or indirectly, the intention of the author(s) communicated in that sentence. It was necessary to infer author intent to exclude citations that were related to study design and methodology but not directly related to the use of hierarchical linear modeling. Citations with the following codes were included in the analysis for research question one: analysis, ICC, justification of selected technique, and explanation or identification of the unit-of-analysis problem. It was possible that a source may have been used for different purposes by different authors. Table 8 summarizes the sources by the (inferred) intention of the author(s). The ICC, a measure of variation across level-2 groups, was
referenced with a citation only two times. This finding did not mean scholars were not reporting the ICC when documenting results of research that use hierarchical linear models. Instead, it should be interpreted to mean, for this sample, scholars did not provide a direct source for the calculation of the ICC in the narrative text of their research. The narrative for research question two includes results concerning ICC reporting in the *HLM Studies* sample.

Table 8. Classification of Citation Use in Text

<table>
<thead>
<tr>
<th>Purpose</th>
<th>n&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>54</td>
</tr>
<tr>
<td>Unit-of-Analysis Problem</td>
<td>40</td>
</tr>
<tr>
<td>Justification of Selected Technique</td>
<td>36</td>
</tr>
<tr>
<td>Intra-class Correlation (ICC)</td>
<td>2</td>
</tr>
</tbody>
</table>

<sup>a</sup>Citations may reflect multiple purposes in text.

Authors cited 36 different sources when writing about the justification for using hierarchical linear models, which represents 32.4% of all sources in this analysis. The finding that almost one-third of citations were used to establish an argument for the use of hierarchical linear modeling with multilevel data may be, in part, attributed to the fact that this sample included articles dating to some of the first applications of hierarchical linear modeling to higher education research. Scholars may have structured their reporting of the method and results to educate readers on both study content (results) and methods (hierarchical linear models). Authors may have believed it necessary and useful to include text *supported by existing literature* to demonstrate why a hierarchical linear model was the most appropriate technique. Descriptions of the unit-of-analysis problem accounted for 36.0% (n = 40) of all sources in the analysis. The unit-of-analysis problem,
which described the statistical issues related to using single-level modeling techniques on multilevel data, was related to the justification of selecting a hierarchical linear model.

The largest percent of citations were linked to the process of conducting the analysis. Fifty-four (48.6%) of all citations were coded as being related to the application of hierarchical linear modeling. Further examination of these sources showed they were also the most frequently cited sources in the HLM Studies sample. These results helped explain the finding of the first analysis, which was a relatively small number of sources served as the key literature when writing about hierarchical linear modeling. It also produced the first key finding of this research: scholars have reached consensus and agreement regarding the sources they cite when writing about the analysis of hierarchical linear models. They consist of four books and the book chapter by Ethington (Bryk & Raudenbush, 1992; Ethington, 1997; Heck & Thomas, 2000; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002). There was, however, no clearly identified list of sources to use when justifying the use of a multilevel model. Differentiating sources between analysis and justification helped to explain the seeming inconclusive results that were reported when examining the list of cited sources based on frequency of use across the articles included in this research.

**Summary for research question one.** The analysis for the first research question consisted of an investigation of both the citations in the text and the sources referenced by each citation. The results demonstrated that the use of citation and sources was more complex than originally believed. Analyzing the sources by frequency of occurrence across articles in the sample proved inconclusive. The results provided
evidence that both supported and contradicted claims about levels of agreement regarding the sources used by scholars of the field. Additional analyses examined the type of sources used and the disciplinary perspectives represented by sources that were journal articles. Scholars relied most heavily on journal sources in the education and psychology fields and were less reliant on articles in statistical fields. The final analysis conducted for this research question interpreted the author(s) intended use of a source in the context of the article text. This analysis led to the key finding that the authors of the articles included in this analysis have reached a clear consensus regarding sources to use when writing about analyzing multilevel data using hierarchical linear models and that is strongly embedded in educational literature.

**Research Question Two: HLM Studies Content Analysis**

Research question two focused on the reporting practices of scholars who have published studies that use hierarchical linear models. As described previously, the increased use of data with a multilevel structure and the applications of theoretical frameworks incorporating constructs or characteristics measured at different levels made it necessary that scholars ensure they have accurately communicated and interpreted the results of all analyses, but particularly those that are new to the field. Using the *HLM Studies* sample from research question one, the analysis for research question two shifted focus from the citations used when reporting the methodology of the research to the narrative description of the methods and results in these articles. It was hypothesized that the results would show that the reporting practices of scholars were inconsistent across articles and, as a result, made it difficult to evaluate the methodological quality of the studies.
**Refinements to method.** The analysis for research question two was based on an analytic framework used by Dedrick et al. (2009, n.d.). Dedrick et al. developed their analytic framework using technical content available in software user manuals and texts on the topic of hierarchical linear modeling. They developed a lengthy coding framework and employed a team of readers. A version of the codebook was shared with the author and served as the basis for the codebook used in this research. The codebook was revised twice during the process of analyzing the articles for this research. First, it was found that several of the codes in Dedrick et al. combined two characteristics into one code. For example, one of the codes reflected whether or not the reader could determine the number of models tested for the study. The set of acceptable codes was a number – to reflect the number of models tested – or qualitative responses “no” or “unable to determine.” Because this research was focused on the presence or absence of certain methodological characteristics, these types of questions were divided into multiple questions, first to determine the presence or absence of a characteristic and second to assess the qualitative element of the characteristic when it was present. This simplified the process of extracting results from the data. A second modification to the codebook at this time included deletion of codes that appears to capture the same information multiple times.

The second revision to the codebook was made to incorporate the desired characteristics that McCoach (2010) listed in her book chapter on reporting studies that used hierarchical linear models. This was done for two reasons. First, although Dedrick et al. (2009) may have applied all codes to their sample, they reported only a subset in
their published article. This made it impossible to discuss the results of this research in
the context of their work, which is one of the only studies similar to this research
identified in searches of the literature. Second, the purpose of McCoach’s book chapter
was to suggest a ‘best practice’ for reporting these hierarchical linear modeling studies.
Questions associated with codes were revised to reflect the criteria put forth by McCoach.
The codes associated with model specification and reporting results were most influenced
by this revision. The practice of revising the codebook reflected the iterative process of
qualitative research (Glesne, 1999). All articles were recoded using the final version of
the codebook.

**Results for research question two.** The hypothesis for research question two
was that scholars were providing insufficient details when reporting results to make it
possible for readers to ascertain the level of credibility of findings. The results presented
here partially support the hypothesis. There was wide variation in how content and
approach to reporting results of hierarchical linear model studies. Two lenses are used to
provide basic interpretations of the results of the analyses conducted for research question
two. The first level is to evaluate the level of agreement within the field regarding the
content that should be included when reporting the results of studies that used
hierarchical linear models. For example, few authors reported information about data
preparation (e.g., outliers or how missing data were handled). This suggested there is
agreement that this is not necessary to include when reporting. In contrast, there appears
to be less agreement regarding how to present information about the theoretical model,
including the specification of relationships among predictors and the variance structures. There were wide variations in how authors reported this information, if at all.

A second lens that can aid the interpretation of these results is how the results compare to recommended content for such studies. In an edited book for reviewers of articles submitted to peer-reviewed journals, McCoach (2010) proposed a structure for the “ideal” content of a study that uses hierarchical linear models. It can be informative to consider in what ways these results align with McCoach’s framework and do not align, independent of the level of agreement within the field. Finally, it may be useful for the reader to know that these results will be discussed in Chapter Five and placed in the context of results from research questions one and three.

**Characteristics of HLM articles.** To understand how higher education scholars are using hierarchical linear models in the field, it can be informative to know what types of modeling techniques are most frequently used in peer-reviewed publications. In this sample, 53 of 60 articles reported results of 2-level nested models. This contrasts with the findings of Dedrick et al. (2009) in their study of hierarchical linear models in general education journals. Dedrick et al. found that modeling growth curves were used more frequently than any other technique. After 2-level nested models, the next most frequently used model was a 3-level, nested model reported in five of 60 articles in the sample. One article reported a 4-level, nested model. The sample included four articles that modeled 2- or 3-level growth curves. Finally, three articles reported results of more than one type of model. In one study the author analyzed the data using a 3-level, nested model and a 3-level, growth curve (Sonnert & Fox, 2012). The purpose of the study was
methodological, intended to determine which modeling technique provided a better fit of the data. Goodness-of-fit tests were used to identify which model yielded a better fit.

Table 9 summarizes the types of multilevel models constructed in the studies included in the HLM Studies sample.

Table 9. HLM Studies by Type of Multilevel Model

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>n&lt;sup&gt;a,b&lt;/sup&gt;</th>
<th>%&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-level nested model</td>
<td>53</td>
<td>84</td>
</tr>
<tr>
<td>3-level nested model</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>2-level nested growth curve</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3-level nested growth curve</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4-level nested model</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

<sup>a</sup> N = 63.  <sup>b</sup>Articles may use more than one model.  <sup>c</sup>Total may not equal 100 due to rounding.

A total of 30 different data sources were represented by the articles included in this research. The majority of studies relied on secondary data analysis, in which authors reanalyzed data collected as part of a national study. Ten articles reported the results of models constructed using original data. The Integrated Postsecondary Education Data System (IPEDS) was used as the most frequent source of level-2 data. Other explicitly mentioned sources of level-2 data included US News and World Report and the Bloomberg Survey.

The NSSE provided data for the largest number of articles. Eleven articles relied on data from the NSSE student survey, two articles used data from the Faculty Survey of Student Engagement (FSSE), and one article used the Community College Survey of Student Engagement. The National Center for Educational Statistics (NCES) served as a source of data for 12 articles and included the Baccalaureate and Beyond (B&B), the National Survey of Postsecondary Faculty (NSOPF), the Beginning Postsecondary
Survey (BPS) and the Survey of Earned Doctorates (SED). Finally, scholars used the full complement of surveys at the Higher Education Research Institute (HERI) in a total of eight articles. Five of eight articles based on analyses of HERI data used more than one HERI survey, reflecting instruments meant to be used longitudinally or to compare results across surveys (Chang, Denson, Sáenz, & Misa, 2006; Cole, 2007; Kim, 2001, 2002; Mayhew, 2012a). Table 10 summarizes the data sources used in at least two studies included in this research. The complete list of data sources is reported in Appendix I.

Table 10. HLM Studies Sample Sources Reported Greater than Two Times

<table>
<thead>
<tr>
<th>Data Source</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Survey of Student Engagement</td>
<td>11</td>
</tr>
<tr>
<td>Original Data</td>
<td>10</td>
</tr>
<tr>
<td>HERI – CIRP</td>
<td>6</td>
</tr>
<tr>
<td>NCES – NSOPF</td>
<td>6</td>
</tr>
<tr>
<td>HERI – Faculty Survey</td>
<td>4</td>
</tr>
<tr>
<td>NCES – Baccalaureate and Beyond</td>
<td>3</td>
</tr>
<tr>
<td>HERI – College Senior Survey</td>
<td>3</td>
</tr>
<tr>
<td>IPEDS</td>
<td>3</td>
</tr>
<tr>
<td>Faculty Survey of Student Engagement</td>
<td>2</td>
</tr>
<tr>
<td>HERI – Spirituality Study</td>
<td>2</td>
</tr>
<tr>
<td>Institutional Course Evaluation Data</td>
<td>2</td>
</tr>
<tr>
<td>NCES – Survey of Earned Doctorates</td>
<td>2</td>
</tr>
<tr>
<td>SUNY – System Study</td>
<td>2</td>
</tr>
</tbody>
</table>

*aSee Appendix I for complete list of data sources.*

The results of these analyses appear to confirm an observation that informed the design of this research: conceptual models such as the IEO model exert a strong influence on the nature of higher education research published in the leading peer-reviewed journals. The structure of theories such as IEO combined with the availability of multi-campus data sets for research make 2- and 3-level nested models a logical analytic approach. This result was not consistent with the distribution of model types identified
by Dedrick et al. (2009), which identified a greater proportion of studies that used growth models. The data sources for the majority of the studies included in the *HLM Studies* sample were based on non- or quasi-experimental designs. It is possible that the studies included in the sample analyzed by Dedrick et al. were based on experimental designs that intentionally incorporated multiple points of data collection. This would explain the higher frequency of growth models reported by Dedrick et al.

**Sample Descriptions.** The format and content of sample descriptions in the articles analyzed for this research showed a great deal of variation. Secondary data analyses (50 of 60 articles) tended to be shorter and provide less description of how the data were collected and the target population. Some referenced another study that provided these details but the majority merely stated what the source of the data were, the number of cases in the data set, and the year the data were collected. More than half of the articles did not specify the sampling procedure (*n* = 33). Of the 27 that reported a sampling strategy 19 used samples of convenience. Both the high rate of non-reported information and the reported high rate of convenience sampling is likely associated with the use of data from existing national surveys such as the NSSE and the CIRP.

Table 11 summarized the sampling strategies used to construct the data sets reported in the *HLM Studies* sample. Descriptions of samples based on original data, however, were more detailed (c.f., Perkins & Craig, 2012; Reason, Cox, Quaye, & Terenzini, 2010; Sonnert & Fox, 2012). McCoach (2010) indicated that details about sampling procedures and the sample should be included in write-ups for studies using hierarchical linear models. While these findings suggest higher education scholars have
reached some agreement regarding how to report information about the sample, it did not align with the recommendations of McCoach.

Table 11. Sampling Strategies Represented in the HLM Studies Sample

<table>
<thead>
<tr>
<th>Sampling Procedure</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unspecified – implicit from data source (i.e., we know how NCES, NSSE collect data)</td>
<td>31</td>
<td>52</td>
</tr>
<tr>
<td>Specified – nonrandom/convenience</td>
<td>19</td>
<td>32</td>
</tr>
<tr>
<td>Specified – random/probability</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Specified – mixed (random at one level)</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Unspecified – Unknown</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

aN = 60. bTotal may not equal 100 due to rounding.

Fifty-six of 60 articles included information about the level-1 units and 54 of 60 articles reported the number of level-2 units. This practice is consistent with what one finds in typical OLS regression studies. A potential drawback to reporting only the total number of level-1 units is that it does not provide a sense of how balanced the distribution of cases is across level-2 groups. Although hierarchical linear models can accommodate unbalanced groups, McCoach (2010) recommended that articles provide sufficient detail to give the reader a sense of how evenly the level-1 cases are distributed across level-2 groups. The results of this analysis suggest higher education scholars prefer a reporting practice more closely aligned with the traditions of an OLS regression study.

Justification of the analytic approach. In her list of desired elements for reporting results of studies that use hierarchical linear models, McCoach (2010) suggested the choice of analytic approach should be “consistent with the purposes of the study and the research questions….“ (p. 125). Fifty-six of the 60 articles included a rationale for using hierarchical linear models and the number of arguments presented
ranged from one to five. The reasons provided clustered into three areas. Thirty-one justifications related to claims that the data in the study had a complex structure. This was most frequently described as “nested” but also included arguments that the data were clustered into groups of differing sizes.

The second theme emerging from the justifications related to the model structure. Twenty-two articles justified the use of hierarchical linear models by stating the modeling process included predictors measured at different levels. Other model-related arguments included ones stating a hierarchical linear model could model cross-level effects and/or allow level-1 coefficients to vary across groups.

The final theme to emerge in the list of justifications relates to statistical or methodological arguments. Eleven articles mentioned that modeling the data using OLS would require one to violate one or more of the assumptions of OLS regression – most frequently that of independence of events. Other justifications included: (a) addressing a unit-of-analysis problem with the data, (b) hierarchical linear models produced more accurate estimates of coefficients, (c) partitioning the variance into level-1 and level-2 components, and (d) reduced Type I error.

A final cluster of responses appeared to fit none of the themes and appeared one time each in an article. These arguments included a statement that student development scholars were not taking advantage of advanced statistical techniques, the study extended the current literature to include multi-campus data, and that a hierarchical linear model was the best technique for measuring growth. Table 12 summarizes the rationales provided by the authors of the articles included in this sample. A comparison of the
justifications for this sample and the regression sample analyzed for research question three is included in Chapter Five of this research.

Table 12. Justification of Hierarchical Linear Model

<table>
<thead>
<tr>
<th>Justification</th>
<th>( n^a )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Structure</td>
<td>31</td>
</tr>
<tr>
<td>Model Structure</td>
<td>30</td>
</tr>
<tr>
<td>Statistical Issue</td>
<td>28</td>
</tr>
<tr>
<td>Other</td>
<td>7</td>
</tr>
<tr>
<td>None Presented</td>
<td>4</td>
</tr>
</tbody>
</table>

\( ^aN = 60. \) \(^b\)Some articles reported multiple adjustments.

**Data preparation and diagnostics.** Data preparation and diagnostic testing is an important first stage for any quantitative analysis (Pedhazer, 1997; Tabachnick & Fidell, 2001). This holds true for multilevel models and is listed as a desired element when reporting results of studies that use hierarchical linear models (McCoach, 2010). The majority of articles did not include information that described how the data were prepared prior to the modeling process. Seven of 60 articles described examination of the data for outliers. Power was mentioned in six articles. In the context of these studies, power was mentioned as established without providing additional detail about how it was tested or established. The exception was an article by Mayhew (2012a) that explicitly mentioned that a power analysis was conducted prior to modeling the outcome.

McCoach (2010) argued for transparency regarding missing data stating the “extent of missing data is clearly reported …and methods for accommodating missing data are described” (p. 124). The authors of the articles included in this research may not agree with the level of transparency called for by McCoach. Scholars reported the extent and management of missing data at substantially higher rates compared to those who
reported content related to power. However, the overall rate of discussing missing data remained lower than 50% of all articles. Twenty-six of 60 articles included narrative regarding missing data. Only one referenced the extent of missing data (Porter, 2006).

Of the 26 that mentioned missing data, 23 described how cases with missing data were addressed. The majority \((n = 14)\) deleted cases with missing data. Six reported a combination of deleting cases and imputing values and substituted the group mean or imputed a value using the EM algorithm \((n = 2)\). Only one explicitly stated the criteria for deleting a case from the data set (Cox, McIntosh, Reason, & Terenzini, 2011). Cox stated that cases with greater than 20% of data missing were dropped from the analysis. Although not always explicitly stated, the majority of imputed values affected level-1 predictors as HLM software drops cases missing level-2 data. Table 13 summarizes results associated with missing data in the sample.

Table 13. Management of Missing Data in HLM Studies Sample

<table>
<thead>
<tr>
<th>Adjustment</th>
<th>(n^{a,b})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deletion (unspecified)</td>
<td>8</td>
</tr>
<tr>
<td>Casewise deletion</td>
<td>5</td>
</tr>
<tr>
<td>Group mean substitution</td>
<td>5</td>
</tr>
<tr>
<td>Drop cases</td>
<td>5</td>
</tr>
<tr>
<td>Listwise deletion</td>
<td>3</td>
</tr>
<tr>
<td>EM algorithm</td>
<td>2</td>
</tr>
<tr>
<td>Unclear</td>
<td>1</td>
</tr>
</tbody>
</table>

\(^a\)\(N = 23.\) \(^b\)Authors may report multiple adjustments.

McCoach (2010) makes several recommendations related to variable selection and preparation in studies using hierarchical linear models. Establishing and reporting reliabilities of measures are a recommended practice when composite variables are included in statistical analyses (Bandalos & Finney, 2010). Of the 60 articles included in
this research, 16 made no mention of reliability. This study did not analyze the types of variables used in the models represented in the sample. As a consequence, it was not possible to determine if the variable list included any composite measures or the authors omitted reliability information from the article. The results suggest reporting information about composite variables has been incorporated into the practices of higher education scholars. This is likely due to the fact that the practice has been recommended for inclusion in any study that uses composite measures. Scholars may be conditioned to include this information based on what has traditionally appeared in other quantitatively oriented articles.

The final aspect of data preparation and diagnostic testing included in the analyses of the articles in this research related to the assumptions of hierarchical linear modeling. Recall that the assumptions for OLS regression and hierarchical linear models are similar. Forty-three articles included no information that suggested, explicitly or implicitly, that the assumptions of hierarchical linear models were tested. Only two articles, by the same author, mentioned three assumptions of hierarchical linear models (Mayhew, 2012b, Mayhew, Vanderlinden, & Kim, 2010). The remaining 15 articles stated that the level-1 residuals were normally distributed or uncorrelated.

**Model construction.** In this research, model construction includes both the specification of the model and how the model is built and tested for fit. Several codes were associated with the model building and testing process. In her chapter for reviewers of hierarchical linear modeling studies, McCoach (2010) stated that the purpose of model information in articles is to facilitate replication of analyses for the purpose of extending
or confirming findings. McCoach suggested that this is best accomplished by reporting
the model as equations, including noting which level-1 slopes are allowed to vary
randomly and which level-1 slopes are modeled using level-2 predictors.

The analysis of article content associated with model specification suggested that
the higher education community has not identified a consistent approach to describing
hierarchical linear models. Using content from the articles included in this research, it
was sometimes difficult to recreate a specified model structure as a set of equations.
Nineteen of 60 articles reported theoretical models using a series of regression equations.
These equations described both the predictors to be tested in the model. More
importantly, reporting a system of regression equations made it possible to determine
whether the intercept or any level-1 slopes were allowed to vary randomly. This
information was represented by the inclusion of symbols representing the between- and
within-group variation. Five articles specified the model using a mixed model equation.
This representation did include information about variance structures. Thirteen articles
included a list of predictors in table form. In three articles there were no meaningful
descriptions of the models. All articles included some narrative, most frequently a
description of variables included in the models.

The selection of predictors to include or test for inclusion in the models was not
consistent across the articles. Almost all the articles (57 of 60) described the variables
selected in the text prior to describing the modeling process. This was interpreted to
mean the predictors were selected based on prior literature or theory. Twenty-one articles
used significance testing to select predictors for inclusion in the model. One used effect
sizes (Umbach & Kuh, 2006) and two made explicit reference to the use of fit statistics such at the AIC or BIC. Other approaches included the use of deviance statistics, running an OLS regression model. Table 14 presents the methods used for selecting predictors to include in models.

Table 14. Justification for Predictor Selection or Retention in HLM Studies Sample

<table>
<thead>
<tr>
<th>Criterion</th>
<th>n^a</th>
<th>%^b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Based on <em>a priori</em> considerations</td>
<td>57</td>
<td>95</td>
</tr>
<tr>
<td>Significance testing</td>
<td>21</td>
<td>35</td>
</tr>
<tr>
<td>Fit statistics (AIC or BIC)</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Selected from OLS regression model</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Deviance statistics</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Limited by degrees of freedom</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Effect sizes</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

^aArticles may include multiple criteria. ^bProportion of total sample.

As described previously, information about variance structures helps the reader understand where to expect variation in the model, particularly variation associated with variables that have significant differences across the groups in the data set. The representation of variance structures occurred much less frequently in the articles in this research. Forty-seven of 60 articles included level-2 predictors in the model. However, the narrative did not indicate whether or not the intercept was allowed to vary randomly. In 11 of 60 articles, it was not possible to determine if the intercept was allowed to vary randomly. In 22 of 60 articles there was clear evidence that at least one level-1 slope was allowed to vary randomly or that a level-1 slope was modeled using level-2 predictors. Fifteen of the 60 articles included in this research did not provide sufficient information to make any determination. If one of the purposes of methods sections of published
studies is for replication, then the current practices in higher education scholarship make it difficult to recreate these models.

Once the theoretical model has been specified, the hierarchical linear model is developed through a series of estimated fixed and random effects, or coefficients. Both Ethington (1997) and McCoach (2010) described a similar modeling process. The models are built in the following sequence: unconditional model, a model with level-1 predictors only, and a model that includes both level-2 and level-1 predictors. Ethington also described a model in which the intercept only is modeled using level-2 predictors. McCoach extended the building process to include a recommendation that each of the models be reported in the results.

Based on the analysis conducted for this research, 44 of 60 articles in the sample included explicit statements about running an unconditional model for the data. Two types of information were used as evidence that the authors built and tested a model that included only level-1 predictors. The first was an explicit statement attesting to the level-1 only model. The second was the presence of results, usually in table form, that showed the coefficients for a model that included only level-1 predictors. In this sample, 28 of 60 articles provided evidence of a level-1 only model, most frequently by listing the coefficients for a model that included only level-1 predictors in a table of results. This was not necessarily the final model of the study.

Finally, 47 of 60 articles modeled the intercept using level-2 predictors. While there are a few examples of level-1 one coefficients that are modeled using level-2
predictors, the analysis of the articles in this sample showed that 22 of 60 articles allowed at least one level-1 predictor other than the intercept to vary randomly.

**Reporting results.** The approaches to reporting results in the articles analyzed for this research revealed a lack of agreement regarding a general approach. It is acknowledged that the research question should have a strong influence on how results are reported, and there are several examples in this research that deviate substantially from both emerging patterns of the field and McCoach’s (2010) recommendations. These examples may be interpreted as reflecting the nature of the research question. However, the majority of studies in this analysis used a similar approach to modeling the outcomes under study. This would suggest that, if there was agreement within the field, we would find evidence of that in the results of this analysis. That was not the case. The results for this section are organized to follow the sequencing of models recommended by both McCoach and Ethington (1997).

A baseline, or unconditional model was reported in 44 of 60 articles. An article was credited with the baseline model if it made explicit reference to the baseline model in text or presented the statistics for the baseline model in table form. Forty-six of 60 articles reported the specific values of the ICC. This statistic represents the proportion of total variance that can be attributed to between-group differences and can inform choices made regarding how to structure relationships among predictors in the theoretical model. Although not tracked explicitly, there appeared to be a tendency by authors in this sample to reference the ICC as a reason to continue with a multilevel modeling analysis.
McCoach (2010) and Ethington (1997) both stated that the next stage of building a hierarchical linear model was to include level-1 predictors only. There are references to this approach in other sources (e.g., Bryk & Raudenbush, 1992) but this appeared as a practical recommendation not one based on statistical need. Twenty-eight of 60 articles reported that a level-1 only model was tested. An article was included in this group if the authors made specific reference to running the model in the text or reported the results in a table of results.

Finally, the model was completed with the addition of any level-2 predictors. Forty-seven of 60 articles modeled the intercept using level-2 predictors. The results made it fairly straightforward to determine which level-2 predictors were included in the models. Authors often described which predictors were significant in the narrative or listed them in table form. It was, however, much more difficult to determine where level-2 predictors were included. Recall that one can model the intercept using level-2 predictors. Conceptually, this is similar to an OLS regression model but statistically corrects for the effects of nested data. It is, however, also possible to model the coefficients (slopes) for the level-1 predictors using level-2 predictors. Fourteen of 60 articles included explicit evidence that level-1 slopes were modeled using level-2 predictors. It was difficult to determine with confidence when level-2 predictors were modeled as cross-level interactions because it was possible that authors modeled the level-1 slopes but did not represent that relationship in the narrative or tested cross-level interactions but excluded them in the final reported model. These results suggest there is
not a strong level of agreement regarding how to organize the reporting of model
coefficients.

The manner in which the predictors are entered into the models influences how
they are interpreted in the results and discussion. McCoach (2010) stated explicitly that
authors should provide this information for the reader. Higher education scholars do not
appear to have reached consensus on the value of this information to the reader. Twenty-
four of 60 articles mentioned the use of grand-mean centering of predictors. Thirteen of
60 mentioned the use of group mean centering. Three articles stated the predictors were
entered as $z$-scores. In 21 articles, it was not possible to ascertain if centering of variables
occurred. Finally, it must be noted that although a combined 37 of 60 articles reported
some centering, authors did not specify whether centering applied to all or a subset of
predictors.

As a final component of how scholars are reporting the results of studies that use
hierarchical linear models, this research analyzed the manner in which fixed and random
effects were reported. The results suggested both a high level of agreement on some
aspects and some lack of consensus in others. Fifty-four of 60 articles reported the
coefficients in table form as a list of estimated effects. There was wide variation in how
that table was formatted. Some reported a simple list of predictors with numerical
coefficients and did not differentiate between level-1 and level-2 predictors. This
approach may reinforce the idea that a hierarchical linear model is equivalent to OLS
regression. Others organized the table to indicate that level-2 predictors were used to
model the intercept only, or intercept and the relevant slopes, which can guide the reader
to understand the relationship among predictors and reinforce the fact that the data are nested, which has implications for findings.

All 60 articles included narratives that identified predictors that were significantly associated with the outcome that was modeled. Nineteen articles included no mention of numerical values for the coefficients in the narrative. Forty-one articles reported the coefficient values in the narrative, typically in parentheses and included the coefficient, standard errors, and/or the level of significance. All articles used an approach that reported only those predictors that were significant and omitted discussion of those that were not significant. Three articles reported the coefficients as a series of regression equations and one reported the findings as a single mixed-model equation. Finally, three of the 60 articles did not report the coefficients (Bowman & Denson, 2012; Cox, McIntosh, Reason, & Terenzini, 2011; Kinzie, Thomas, Palmer, Umbach, & Kuh, 2007). This result can most likely be attributed to either the journal in which the article was published or the nature of the research question. Table 15 summarizes the methods of reporting fixed and random effects represented in the HLM Studies sample.

Table 15. Methods of Reporting Fixed and Random Effects in HLM Studies Sample

<table>
<thead>
<tr>
<th>Reporting Method</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of estimated effects</td>
<td>54</td>
<td>90</td>
</tr>
<tr>
<td>Verbal description</td>
<td>41</td>
<td>68</td>
</tr>
<tr>
<td>Not reported</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Series of regression equations</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Single mixed model equation</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

*aSome articles included multiple methods of reporting fixed and random effects. bProportion of total sample.

The ability to partition variance into between- and within-groups components is a key benefit of hierarchical linear models. Thirty-three of 60 articles listed the estimated
values of the variance parameters in a table. Listing the variance in a table did not guarantee that the variance was discussed in a meaningful way. Only 22 of 60 articles included a mention of variance in the narrative. Six articles incorporated variance structures into the representation of the theoretical models as a mixed model equation or system of equations. Finally, 11 of the 60 articles made no mention of variance structures at all.

The lack of agreement suggests higher education scholars have not reached consensus about how to report the fixed and random effects in studies using hierarchical linear models. What this means for the field is less clear. Reporting lists of coefficients without representing relationships between the predictors or across groups may leave readers with the impression that hierarchical linear models are the “same” as an OLS regression model. While they have some similarities, there are key differences in the statistical approaches that are relevant to both how we use them and how we interpret results. It is also worth noting that the cumulative practices represented by the results are not consistent with the recommendations of McCoach (2010).

The omission of information about the variance structures may also reinforce the misconception that hierarchical linear models are almost identical to OLS regression models. It is important to acknowledge that the majority of these studies were published prior to McCoach’s (2010) framework. The implications of this are discussed in Chapter Five. Finally, given the frequency with which Ethington’s 1997 work was cited in these articles, it was not surprising that when authors described a process for modeling the outcome of their study, they often followed the approach described by Ethington and
others. This may explain why details about the model building process were not described in the majority of studies.

**Tools for constructing the model.** Dedrick et al. (2009) reported the frequency with which authors reported the software package used to build the models. In addition, estimation algorithm and methods affect the estimation of fixed and random effects. Coding conducted on this data set showed that 37 of 60 articles included at least the name of the software package used in the analyses and 23 did not report the software used. Software represented by the articles include: HLM, MlWin, SAS, and SPSS. Of the 37 articles that reported the package used, 28 used a version of HLM. The version ranged from HLM 2L to HLM 6.02. With the exception of a single study published in 2011, studies that reported use of an early version of HLM (versions 2L through 5.05) were published prior to 2006 suggesting an association between version and publication year.

Given that the citation analysis of research question one showed that higher education scholars relied on publications by Bryk and Raudenbush (c.f., 1992, 2002), it was not unexpected to find that higher education scholars also used software developed by the same authors. An even smaller number of articles included information associated with the estimation methods/algorithms used which can be manipulated in software applications such as HLM. Six of 60 articles reported either the estimation method or algorithm (c.f., Cox, McIntosh, Reason, & Terenzini, 2011; Hox, 2000; Jessup-Anger, 2012; Smeby & Try, 2005). While it was reasonable to interpret this finding to mean that scholars used the default settings for the software packages, readers cannot determine with certainty what settings were used from the narratives in the sample.
Summary for research question two. The results of research question two show that higher education scholars are not in complete agreement regarding how to structure content when reporting results of studies that used hierarchical linear models. This inconsistency may support interpretations that explain why scholars continue to use OLS regression models on multilevel data. The results will be discussed in the context of the results from research questions one and three as well as the implications in Chapter Five.

Research Question Three: Regression Studies Content Analysis

Research question three explored the ways in which scholars interpret hierarchical linear modeling concepts in the context of studies that applied single-level modeling techniques on data with a complex structure. The question that framed this analysis asked what reasons do scholars give in published articles to justify the use of single-level modeling approaches on complex data? It was hypothesized that there would be a lack of consensus regarding the rationales scholars reported in published studies that used single-level regression modeling techniques to model data with a multilevel structure or variables measured at different levels.

This question applied content analytic techniques to identify text that related to data structure, justification of analytic technique, and adjustments or corrections to modeling procedures. The Regression Studies sample, described earlier in the chapter, served as the source of articles analyzed in this research. Because the general focus of this research was hierarchical linear modeling, this sample consisted of published articles meeting the following criteria: (a) the outcome measure was continuous, (b) the data source had a clearly defined complex structure with at least 10 level-2 units, and (c) the
data were modeled using a single-level regression model. A total of 50 articles met all conditions and were included in the analysis.

**Refinements to method.** The analytic plan for this research consisted of identifying and coding text segments that related to the management of multilevel data in a single-level model. The codebook and form (Appendix E) were used as the guide for coding. Several modifications to the codebook were made throughout the coding process. The first version of the codebook included coding to indicate the absence or presence of characteristics such as stating or acknowledging a data set had a complex structure. A yes/no code did not represent the full spectrum of information that informed that response so a series of refinements and new codes were introduced to identify different forms authors might have used to characterize a data set’s structure.

A second refinement to the codebook included the addition of codes to identify the level of measurement for predictors included in the models for these articles. The decision to include this code was supported by Hox (2010) who described statistical implications related to data structure and aggregated/disaggregated variables to a single level. All articles were reanalyzed to assign a value to each article for this code. These refinements reflect the challenges of using content analysis for exploratory research. The original coding for justifications and adjustments was not inclusive of all evidence identified in the text during analysis. The final codebook reflects the complete list and all articles were recoded using the amended codebook. The results presented here reflect the final code determination.
**Results for research question three.** The results of the analysis for research question three are organized into three areas. First, coding related to whether and how authors acknowledged that the data or model had a multilevel structure is presented. Next, justification patterns and themes are explored. Finally, the analysis examined statistical adjustments scholars made to “correct” violations of statistical assumptions and/or procedures.

**Data structure and model specification.** One of the choices a scholar makes when conducting research is to select the most appropriate analytic technique for the research question and the data to be analyzed. While it was likely not reasonable to assume that scholars communicate all such considerations and decision in the reporting of their research, coding was conducted to determine the frequency with which an author acknowledged that the data had a complex structure. This was coded using direct evidence of the structure, such as the words “multilevel,” “cluster,” or “nested.” Twelve articles in the *Regression Studies* sample included an explicit statement acknowledging that the data were multilevel. Multilevel structure was acknowledged indirectly by reporting the number of groups. Forty-one of 50 articles reported the number of groups represented in the data set. One interpretation of these results is that scholars are assuming the reader infers that by reporting the sample sizes for each level of the data set means the data have complex structure.

Given that hierarchical linear modeling could be considered an emerging statistical technique in higher education research during the span of years included in the samples for this research, it was also informative to examine results using the lens of year
of publication. The first explicit mention of data having a multilevel structure did not appear until 2003 in an article by Perna. The results did seem to suggest a “tipping point” regarding acknowledgement of complex data structure. Of the 13 articles published between 2010 and 2012, 12 made explicit reference to data having a multilevel structure. These 13 articles represent 54.2% of all articles that acknowledged a complex data structure.

Higher education scholars have long attended to the effect of institutional context in their research, and widely adopted theories exist that include context in their conceptual frameworks (Astin, 1991, 1993; Pascarella, 1985; Pascarella & Terenzini, 1991, 2005). Modeling outcomes using these frameworks makes it necessary to include predictors that are measured at different levels, which can produce erroneous results (Hox, 2009). The articles in the regression sample were coded to identify those that made specific mention of a model structure. Six of the 50 articles included explicit statements that the specified model included both level-1 and level-2 predictors. To further explore the nature of the models, each article was coded to identify the presence of level-1 and level-2 predictors in the specified models. Seventeen articles in the Regression Studies sample included predictors measured only at level-1. As will be discussed in Chapter Five, models with predictors measured at only one level do not always require the use of multilevel techniques. However, 33 of 50 articles in the Regression Studies sample listed predictors at different levels, which is associated with multilevel techniques such as hierarchical linear models. Table 16 summarizes the level of predictors represented in the models of the Regression Studies sample.
Table 16. Levels of Predictors in Regression Studies Sample

<table>
<thead>
<tr>
<th>Predictor Level</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level-1 Only</td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td>Level-2 Only</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Level-1 and Level-2</td>
<td>33</td>
<td>66</td>
</tr>
</tbody>
</table>

N = 50. Proportion of total sample.

Justification of a single-level regression model. The primary focus of research question three was to identify if and how authors justified using a single-level model with multilevel data. Twenty-four of 50 articles included a rationale for the use of a single-level model. The earliest article with justification is Kuh’s and Hu’s (2001) article published in the Review of Higher Education.

Next, the data were analyzed to identify the argument used to justify the use of a single-level regression model. Authors provided from one to four arguments to justify the selection of a single-level regression model (Table 17). Of the 24 articles that included some form of justification or rationale for using regression or against using a multilevel model, 14 included a single argument. Two articles provided four arguments. One person served as lead author on each of these articles.

Table 17. Articles Providing Justification for Analytic Technique in Regression Studies Sample

<table>
<thead>
<tr>
<th>Number of Arguments</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>26</td>
<td>52</td>
</tr>
<tr>
<td>One</td>
<td>14</td>
<td>28</td>
</tr>
<tr>
<td>Two</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Three</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Four</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

N = 50. Proportion of total sample.

The arguments provided included 14 distinct arguments that clustered into three themes. These themes are: (a) regression and hierarchical linear modeling produce
equivalent results; (b) arguments related to the composition of predictors in the model; and (c) arguments related to the variation across groups. Four arguments, each used one time, did not fit into any of the above themes. Table 18 summarizes the arguments identified in the analyses of justification.

Table 18. Arguments Presented to Justify Single-Level Model in Regression Studies Sample

<table>
<thead>
<tr>
<th>Argument</th>
<th>(n^a)</th>
<th>Theme(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model is single-level</td>
<td>6</td>
<td>II</td>
</tr>
<tr>
<td>Inclusion of level-2 predictors not supported by prior research</td>
<td>5</td>
<td>II</td>
</tr>
<tr>
<td>Insufficient variation across groups (ICC reported)</td>
<td>5</td>
<td>III</td>
</tr>
<tr>
<td>Regression and HLM produce similar results (cited)</td>
<td>4</td>
<td>I</td>
</tr>
<tr>
<td>Wanted to follow changes in beta coefficients</td>
<td>4</td>
<td>II</td>
</tr>
<tr>
<td>Ran both analyses and results were similar</td>
<td>4</td>
<td>I</td>
</tr>
<tr>
<td>Regression and HLM produce similar results</td>
<td>2</td>
<td>I</td>
</tr>
<tr>
<td>Insufficient variation across groups (ICC not reported)</td>
<td>2</td>
<td>III</td>
</tr>
<tr>
<td>Regression results are easier to interpret</td>
<td>2</td>
<td>I</td>
</tr>
<tr>
<td>Two few level-2 groups</td>
<td>1</td>
<td>IV</td>
</tr>
<tr>
<td>Degrees of freedom reduced in an HLM model</td>
<td>1</td>
<td>IV</td>
</tr>
<tr>
<td>Data cannot be disaggregated</td>
<td>1</td>
<td>IV</td>
</tr>
<tr>
<td>Similar studies used regression</td>
<td>1</td>
<td>I</td>
</tr>
<tr>
<td>Focused on level-1 effects</td>
<td>1</td>
<td>II</td>
</tr>
<tr>
<td>Software can correct for clustering problems</td>
<td>1</td>
<td>IV</td>
</tr>
</tbody>
</table>

\(^a\)Articles included multiple arguments. \(^b\)I = Regression and hierarchical linear models are equivalent; II = Predictor composition; III = Group variation; IV = Other.

The theme reported most frequently as a justification for regression was that regression was equivalent to hierarchical linear modeling, and therefore could be treated as interchangeable. Arguments related to this theme occurred 13 times. The authors of four articles argued that regression and hierarchical linear modeling were equivalent and included a citation to support this claim. The sources cited were Ethington’s 1997 book chapter and the 2009 article by Astin and Denson, both summarized in Chapter Two.

Two articles included arguments that regression and hierarchical linear models were similar without a citation to support the argument. Four articles included a statement that
the authors had analyzed the data using regression and hierarchical linear modeling and that the results were the same (Park, 2009; Steinberg, Piraino, & Haverman, 2009) or similar (Toutkoushian & Smart, 2001; Zhang, 2005). Two articles included the justification that regression was used because it was easier to interpret than the results of a hierarchical linear model. One article included the claim that using regression was consistent with prior research, perhaps implying that it would be easier to contextualize findings.

The second theme related to the composition of the model. Six articles included a statement that the model for the study included predictors measured at only one level. It is worth noting that one of the articles that included this argument had predictors measured at both level-1 and level-2. This inconsistency was neither noticed by the authors nor through the peer-review process prior to publication. Five articles made statements that the inclusion of level-2 predictors was not justified by prior research. Two of these articles were written by the same lead author and included two citations to support the argument. The authors of four articles wanted to follow changes in beta-coefficients and claimed that this was not possible in a hierarchical linear model. At least one author on each of these articles is affiliated with the same institution, which suggests this argument may be rooted in an organizational philosophy regarding the modeling techniques. Finally, one article included the argument that they were interested only in level-1 effects, therefore regression was appropriate.

The third theme of the justifications for regression related to the amount of variation in the outcome variable. Most descriptions of the sequence of procedures to
build a hierarchical linear model began with a calculation of the ICC (Hox, 2009). Described in the review of the literature, the ICC reports the proportion of total variation in the outcome measure that can be attributed to differences across groups. A total of seven articles included a statement related to the ICC to justify the use of regression with multilevel data. Five of the articles that included mention of the ICC in the literature reported a calculated value of ICC and reported that there was too little variation across groups, which justified the use of regression instead of a hierarchical linear model. Not all authors specified what “too little variation” meant. The threshold reported ranged from 2% to 10%. Two of the articles state there was “evidence of limited clustering” or that the ICC was low but did not report a value for the ICC.

The remaining arguments were reported one time each. One article claimed there were too few groups in the data set to require the use of hierarchical linear models. While statisticians have shown that, *under very specific conditions*, one can use single-level models on clustered data (Maas & Hox, 2005), I was not able to determine from the narrative from this article if those conditions were met. The article that included this justification described 18 level-2 groups, which would suggest this is not a valid justification for regression. The implications of level-2 group size are subtle and perhaps missed by authors and peer-reviewers. However, these differences can lead to increased risk of erroneous results.

One author claimed that the effects of clustering could be corrected using statistical software. This is true and is discussed in the next section of results. The data for one study was a secondary analysis from multiple institutions, which makes the data
clustered. However, the original data collection did not assign an institutional identifier to each case, reducing the data set to a single-level sample. Finally, one article included a comment that the use of hierarchical linear models reduced the degrees of freedom, which limited the number of predictors one can include in the model. While this is true, the degrees of freedom are similarly reduced in a regression model because the data were clustered within groups not because one was using hierarchical linear models.

**Corrections to the model.** The final analysis for this research question examined the adjustments or corrections to single-level regression procedures made by scholars when analyzing multilevel data using a single-level regression model. This was conducted because the review of the literature produced some evidence that some software packages include procedures to produce robust standard errors that may produce statistically accurate results when analyzing nested data. For example, the `svy` command in Stata is intended to correctly analyze data with a complex structure. Corrections, or adjustments, to the modeling process were not always included as justifications for using regression but did appear to perhaps correct both reported and unreported limitations to OLS regression methods. Fifteen articles in the regression article sample included at least one adjustment. Twelve articles included a single correction to the regression model. Two articles included a description of two corrections and one article included a list of four adjustments to the modeling process. Tables 19 and 20 summarize the results of these analyses.
Table 19. Adjustments to Model Building Process in Regression Studies Sample

<table>
<thead>
<tr>
<th>Number of Adjustments</th>
<th>n(^a)</th>
<th>(%)^b</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>35</td>
<td>70</td>
</tr>
<tr>
<td>One</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>Two</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Four</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

\(^aN = 50. \ ^bProportion of total sample.

Table 20. Distribution of Adjustments to Modeling Process in Regression Studies Sample

<table>
<thead>
<tr>
<th>Adjustment</th>
<th>n(^a, b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used a more conservative (p)-value</td>
<td>8</td>
</tr>
<tr>
<td>Used Stata software</td>
<td>5</td>
</tr>
<tr>
<td>Used AM software</td>
<td>2</td>
</tr>
<tr>
<td>Adjusted for standard errors (non-specific)</td>
<td>1</td>
</tr>
<tr>
<td>Reduced number of model predictors</td>
<td>1</td>
</tr>
<tr>
<td>Transformed level-1 predictors to accommodate reduced degrees of freedom</td>
<td>1</td>
</tr>
<tr>
<td>Modeled only level-1 predictors</td>
<td>1</td>
</tr>
<tr>
<td>Used weighting</td>
<td>1</td>
</tr>
</tbody>
</table>

\(^aN = 15. \ ^bAuthors may report multiple adjustments.

Eight of the 15 articles that included a statistical correction listed the use of more conservative \(p\)-values as an acceptable correction and adjustment. Seven of the eight articles specified the \(p\)-values, which ranged from .01 to .001. One article did not specify a precise value for \(p\). Five articles used the software package Stata, and its \texttt{svy} procedure to produce adjusted standard errors. Two articles, with the same lead author, used another package called AM Software, published by the American Institutes for Research and designed to analyze complex data sets. One article reported a correction to the standard errors but did not specify how that was accomplished.

Two articles made adjustments to the model to correct for the effect of complex data structure. The authors of two articles reported that the number of predictors was reduced. One article included no additional detail, but the second of the two made explicit reference to the fact that the degrees of freedom were reduced because the data
have a complex structure. The authors of one article also stated that level-2 predictors were excluded from the model. This correction occurred in tandem with the use of Stata statistical software.

Weights were also used for reasons unrelated to multilevel data structures and were reported in the text of 16 articles. However, only one author made specific mention of weighting as an appropriate adjustment for multilevel data. Finally, one article mentioned the use of dummy coding to address the issues related to complex data structure. It was not clear from the review of the literature that dummy coding adequately addresses the issues of variables at different levels in a single-level model (Kuh & Hu, 2001).

**Summary for research question three.** The third question in this research explored the reasons scholars report in published articles for not using multilevel models, specifically hierarchical models, in research using data sets with complex structure. The analysis for this research question used an exploratory content analysis that identified knowledge claims that provide a framework for making inferences about the choices scholars make when conducting this type of research. The results of the analyses on this sample showed that authors did not explicitly acknowledge that the data in their research had a complex structure. A clear majority, however, reported the number of level-2 and level-1 units in their narratives. An even smaller percentage acknowledged that the predictors in their models consisted of variables measured at different levels. Although a minority of authors make explicit justifications for the use of a single-level model, the rationales provided by those who did justify their analytic choice coalesced into four
themes: (a) the equivalence of HLM and regression results; (b) too little between-group variation, (c) predictors measured at only one level, and (d) prior research did not identify level-2 predictors as significant. Finally, the analysis showed that a minority of scholars made corrections to single-level statistical procedures. The most common correction was to use a software package designed to compensate for the statistical effect of clustered data.

Chapter Summary

This chapter provided a detailed summary of results of this research. The chapter described the two samples analyzed for this research, described the research questions and hypotheses. Because each research question required the use of a different analytic approach and/or sample, the results were presented by research question. Findings showed that the scholars relied on a relatively small set of sources when referencing materials related to methodology and often omitted details about data preparation, diagnostic testing and specification of model and variance structures. In studies that applied regression techniques to data with a complex structure, scholars have not consistently addressed the methodological issues associated with this analytic choice and did not consistently present accurate arguments or corrections for the analyses. The next chapter will synthesize results across analyses, discuss the implications for scholarly practice, future research, and examine the limitations of the research.
CHAPTER FIVE
SUMMARY AND DISCUSSION

Guided by three questions, this research has created a foundation for understanding the practices scholars used when reporting the results of studies analyzing data with a multilevel structure. Chapter Five summarizes the general findings of this research and interprets them in the context of the literature and practices of the field. The chapter begins with a review of the purpose of the study, the literature that informed the study, the design and analytic plan. Following this summary, the results are discussed in the context of related literature, the implications of the findings, and recommendations for future research.

The research presented here makes several contributions to the scholarship of the field. Content analyses using scholarly publications as the source of data are not well represented in the higher education literature (c.f., Cheslock & Rios-Aguilar, 2011; Ferron et al. 2008). Examination of cited sources provided insight into the resources deemed important to our collective understanding of hierarchical linear models. Examining the methodological content of articles that reported the results of hierarchical linear modeling provides information about what authors identify as essential information related to the technique. Even more informative, perhaps, is what was not reported. Finally, understanding the arguments authors make for not using hierarchical linear modeling with multilevel data identified gaps in the field’s understanding of the
application of the technique. Together the analyses of this research describe the shared understanding of hierarchical linear models and its place as an analytic tool.

**Relevant Literature**

The review of the literature of the field identified the need for advanced statistical tools to conduct research on topics of interest to higher education scholars (Ethington, 1997; Pascarella & Terenzini, 2005). Astin’s IEO framework, for example, has been influential in shaping the conversation in higher education about the impact of the college experience and has been applied to a variety of outcomes such as institutional commitment and intellectual self-concept (Cole, 2007; Strauss & Volkwein, 2004). Application of such frameworks to the study of higher education problems makes it necessary to include predictors measured at different levels. However, constructing single-level models using predictors measured at different levels can lead to misestimated standard errors and increased risk of Type I errors in inferential tests (Hox, 1995). Some of these errors can be corrected using modified regression procedures, but it was not clear from the review of the literature that these practices had been applied widely to studies in the field.

Perhaps associated with the emerging importance of complex theoretical/conceptual frameworks such as Astin’s IEO model, the first decade of the 21st century has also produced an increase in the number of national studies of college students and/or faculty. These new studies include the NSSE and its variations, the MSL, the Wabash Study, and the Student Experience in the Research University survey (SERU). These data sets make it easier to incorporate predictors measured at the level of the institution
into models of college outcomes. However, clustering of individuals within groups violates fundamental assumptions of OLS regression modeling. Failing to account for and/or test for statistical differences among groups on outcome measures may also lead to increased risk of Type I errors and the identification of predictors as significant when they are not.

**Statement of the Problem**

Higher education researchers draw from a variety of epistemic and disciplinary perspectives and consider a broad spectrum of methodologies to be valid approaches to constructing the knowledge of the field (Dressel & Marcus, 1982). Articles published in the leading field-related journals serve as evidence of this diversity of content and method. Recently, multi-campus surveys have been the target of strong critiques (c.f., Porter, 2011; Porter, Rumann, & Pontius, 2011). The focus of these critiques, however, has been on the validity of self-reported data, not the tools of analysis.

Prior to the development of multilevel modeling techniques, researchers wishing to model outcomes using multilevel data used OLS regression or its appropriate variant. They did so knowing that key statistical assumptions for OLS regression were not met. The assumption of independence is violated when constructing data sets that consist of cases nested within some group. Under certain circumstances, this violation can be managed statistically in an OLS regression model (Heeringa, West, & Berglund, 2010). Additional issues arose when the proposed model included predictors measured at different levels. For example, a study may seek to model an outcome measured at the level of an individual student using predictors measured at the level of the student and the
level of the institution. Multilevel techniques, including hierarchical linear modeling, accommodate the multilevel structure of data and provide estimates that are less prone to error.

Multilevel techniques resolve both the issue of nested data and make it possible to model outcomes using predictors measured at different levels. One might then expect the scholars of the field to substitute OLS regression with multilevel modeling and that, over time we would observe a downward trend in the number of published studies that used OLS regression in the journals represented by this research. Scans of the peer-reviewed literature, however, indicated that higher education scholars have continued to use single-level regression techniques.

Based on preliminary scans of the literature, it was not unexpected there would be little direct evidence to explain why scholars continued to use single-level regression models. Scholarly articles and essays that evaluate why scholars make their methodological and analytic choices do not appear in the journals of the field. This is likely attributable to the specified focus of the journal as represented by editorial guidelines. If such articles are published in statistical journals, their impact is necessarily limited as higher education scholars may not review the content of such journals as an ongoing practice. In Chapter One, I suggested that it was possible the field, as a community of scholars, did not have sufficient understanding of multilevel modeling to facilitate the transition from OLS regression to a more statistically appropriate multilevel analytic approach as the preferred analytic method for modeling outcomes on multilevel data or when including predictors measured at different levels in the same model. The
results of this research suggest that journal guidelines, the peer review process, and historical practices of journals may play a more significant role than scholar expertise on the choices scholars make while conducting their research. This concept is explored in more detail in a subsequent section of this chapter.

Review of Methods

The current study applied content analytic techniques to two samples constructed from articles published in four journals with an explicit focus on reporting higher education research: Journal of College Student Development, Journal of Higher Education, Research in Higher Education, and the Review of Higher Education. These journals were selected because they were identified as having the highest rate of citation in the field and the content represented the breadth of topics and methodologies of interest to higher education scholars (Tight, 2008). It seemed reasonable that the articles of these journals would be representative of the highest methodological and analytic quality and were, therefore, suitable for use as sources of data for this research.

Two samples of articles were required for this research. The first sample, described as HLM Studies sample, consisted of articles published between 2000 and June 30, 2012 that used hierarchical linear modeling as a primary method of analysis. This sample was used in the analysis for research questions one and two. The HLM Studies sample consisted of 60 articles, of which 55% \( (n = 33) \) were published in the journal Research in Higher Education. The second sample, termed Regression Studies sample, included articles that applied single-level regression techniques to multilevel data. The
Regression Studies sample included 50 articles with Research in Higher Education, again, contributing the greatest number of articles \((n = 17)\).

Three questions guided the analysis for this research. The first question was addressed using a citation analysis of the HLM Studies sample. Descriptive statistics provided insights into citation patterns scholars used relating specifically to reporting research methods, analyses, and methods related discussion of results and implications. The second question identified patterns in reporting practices related to studies using hierarchical linear modeling. Content analysis using an adapted checklist developed by Dedrick et al. (n.d.) addressed this research question. The final question in this research was addressed using the Regression Studies sample. This analysis used an exploratory content analysis to identify reporting patterns that were used to justify single-level regression on a complex data set. Again, descriptive statistics were used to identify general themes across all articles in the Regression Studies sample.

**Summary of Results**

The first research question asked what methodological sources have been cited by higher education scholars who have published studies that used hierarchical linear models? It was hypothesized that scholars were relying on a relatively narrow collection of sources. The analyses related to this research question supported the hypothesis. The second research question asked using Dedrick et al.’s (2009) analytic framework for examining the narrative content of studies using hierarchical linear modeling, what methodological issues were included/omitted in narratives of studies using the technique? It was hypothesized that reporting practices for this technique would not be consistent
across articles. The analyses for this question partially fulfilled the hypothesis. The final question of this research was *what reasons did scholars give in articles to justify the use of single-level modeling approaches on data with a complex structure?* The hypothesis for this question argued that the analyses would suggest there is wide variation in the patterns of justification. The results partially supported this hypothesis.

**HLM Studies Sample Citation Analysis**

The analyses for research question one identified a total of 113 different sources across 60 articles included in the *HLM Studies* sample. The most frequently cited source represented two editions of a single text: Bryk and Raudenbush’s 1992 text on hierarchical linear modeling and the second edition published in 2002. The 2002 publication was cited in 56.7% articles in the sample. The first edition of this text, published in 1992, was cited in 28.3% articles. Combined, these books were cited in 85.0% of the sample articles analyzed. Only one other source was cited in more than 10 articles, a 1997 book chapter by Ethington. This chapter, from *Higher Education: Handbook of Theory and Research* was cited in 33.3% articles. Seventy-three percent of sources were cited exactly one time across all 60 articles of the sample. The results suggested scholars have generally agreed on the “canon” for hierarchical linear models. The impact of these sources was further reinforced by the analysis conducted for research question two.

The type of sources used by scholars was also analyzed. The collection of 113 cited sources represented 38 books or book chapters, 13 technical manuals, and 52 journal articles. Further analyses of the journal articles showed that they represented 33
different journal titles. The majority \( (n = 18) \) of these journals had a focus on education or educational psychology. Five of the 33 journals were identified as having a statistical or methodological focus and the remaining 10 represented journals in management, sociology, or economics.

Finally, it was necessary to acknowledge that the use of citation is in the context of the narrative. The final analysis for research question one consisted of a review of each citation \textit{in situ} in the narrative to infer the purpose of the citation. Fifty-four sources related to the method of analysis. Thirty-six were in reference to the justification of multilevel analysis, and 40 of the 113 articles were cited in the context of describing the problem of selecting the appropriate unit of analysis.

**HLM Studies Sample Content Analysis**

The second question of this research focused on how scholars framed and described the analyses and results of studies that used hierarchical linear models. Each article was analyzed using a coding framework adapted from a study conducted by Dedrick et al. (n.d., 2009). This framework described four broad areas of narrative content related to method and analysis in studies using hierarchical linear models: (a) data sources and selection of analytic technique; (b) data preparation; (c) model construction; and (d) presentation of results. Results from the analyses for this research question suggest the hypothesis was partially fulfilled.

**Data and selection of analytic technique.** Results showed that higher education scholars were using hierarchical linear models to construct two-level models and that in the majority of cases these models consisted of faculty or students nested within
institutions. National surveys such as the NSSE and the studies conducted by HERI served as a data source for the majority of studies included in this research. Nearly 50% of articles did not include information about how the sample was constructed for the study, perhaps based on an assumption that readers would know this information based on the data source. Ninety-three percent of articles included some language that justified the use of a hierarchical linear model \((n = 56)\). Justifications represented three themes: (a) data set had a multilevel structure; (b) predictors were measured at different levels, and (c) one or more of the assumptions of regression would be violated. While one might argue the rationales are sufficiently similar to represent the same general concept, each presented a slightly different point of view regarding the reasons one might use hierarchical linear model.

**Data preparation.** The results showed that higher education scholars have not, as general practice, adopted a practice to include information about data preparation in the narrative of their articles. Scholars included little content that described processes for preparing the data. Only 11% of articles included content that verified outlier analysis. Ten percent of articles included a discussion of power. Fewer than 50% of articles included any mention of missing data and narrative tended to focus on how missing data were managed. The majority simply deleted cases and only one article described the criteria for deletion. Computational transformations to replace missing data included replacement with a group level mean or imputed value via the EM algorithm. No author specified precisely which variables had missing data. While these results suggest higher
education scholars are in general agreement regarding what content to include regarding data preparation, it is not consistent with the standards proposed by McCoach (2010).

**Model construction.** The coding associated with model construction examined two aspects of constructing a hierarchical linear model. The first set of codes described how variables were selected for inclusion and/or retention in a model. The second set of codes described the process scholars used for constructing the model. The results suggested that higher education scholars ground the selection of predictors in prior literature, informed by theoretical and conceptual frameworks, as 95% of articles made explicit reference to a priori variable selection. Thirty-five percent of articles retained variables based on statistical significance. Other approaches were used (e.g., fit and deviance statistics, effect sizes) but were mentioned in no more than two articles in the sample. The variables selected represented data measured at different levels and would be expected given the focus of this research. Coding from these analyses revealed that scholars did not report explicit information about specification of variance structures.

While 78% of articles included information about level-2 predictors, scholars generally omitted information that made it possible to determine when the intercept and level-1 variables were allowed to vary randomly and or modeled using level-2 predictors.

As described in Chapter Two, Ethington (1997) presented an approach to building a model that included testing the unconditional model, a model that included level-1 predictors only, and a “full” model in which level-1 predictors, including the intercept, were modeled using level-2 predictors. The coding from this research suggested that scholars have adopted this approach to model building. Seventy-three percent of articles
reported the unconditional model. Forty-seven percent of articles provided evidence of a model that included level-1 predictors only and 78% of articles provided clear evidence that the intercept was modeled using level-2 predictors. Finally, 37% of articles included evidence that a level-1 slope other than the intercept was allowed to vary randomly. It is important to note that, although the approach authors used to report models aligned with the one described by Ethington, it was frequently not possible to rebuild the actual processes used to construct the completed model from the narrative content.

**Presentation of results.** The coding from this research suggested that there is no generally accepted practice for reporting results. Ninety percent of articles reported coefficients in a table either in the main narrative or as an appendix. However, the format for these tables showed a great deal of variation. The most common practice appeared to be a format that was similar to OLS regression reporting. Scholars did not always differentiate between levels of variables, leaving the reader to infer that only the intercept was modeled using level-2 predictors. Variance parameters were included in tables in 55% of articles.

A second aspect of reporting is the manner in which authors incorporated results in narrative. Thirty-two percent of articles described the results without inclusion of numerical values in narrative. Instead, these scholars used a qualitative description of predictors with statistical significance and omitted numerical values completely. Thirty-seven percent of articles included language related to the variance components in the narrative and 18% of articles included no mention of variance structures in any form. The coding associated with reporting coefficients and variance structures suggested that
higher education scholars have not reached consensus about how to report fixed and random effects in studies that used hierarchical linear models. The implications for these results are discussed in subsequent sections of this chapter.

The hypothesis for this research question was that there would be variation in reporting practices and that this variation would make it difficult to ascertain if scholars were conducting these analyses with methodological accuracy. Do we get it right? The results partially fulfilled the hypothesis. The descriptive statistics here suggest that scholars have not reached consensus on the essential elements when reporting the results of studies that used hierarchical linear models. In that respect the hypothesis is fulfilled. However, the evidence is not sufficiently conclusive to support a claim that the field lacks understanding of the technique. There are some articles in the sample that seem to reflect insufficient knowledge on the use of the technique. However, for each article in the sample that might be used as evidence that scholars do not understand the analytic technique, one could also identify an article that could be used as an example of high quality reporting of the method and results.

Regression Studies Sample Content Analysis

The third question of this research examined a specific aspect of reporting practices found in studies that used OLS regression models. The research question that informed this content analysis was what reasons did scholars give in published articles to justify the use of single-level modeling approaches on data with a complex structure? These results are organized into three areas.
The first set of codes for this analysis related to the data structure and model specification. Eighty-two percent of articles in the sample included language that acknowledged, directly or indirectly, the data set for the study had a complex structure. As described in Chapter Two, the predictors selected for inclusion in a model can be used to justify the use of a single-level regression model. To test whether or not the specified model might justify the use of OLS regression, the levels of the predictors included in the regression models were analyzed. Results showed that 34% of models included predictors measures at only one level, and this was always at the lowest level of the model (e.g., at the level of the individual student). Sixty-six percent of specified models included predictors measured at different levels.

A second set of codes for this research explored the original focus of the research question, which were the reasons higher education scholars gave to justify the use of OLS regression, a single-level model, with multilevel data. Forty-eight percent of articles included some explicit statement to justify the use of a single-level model. Authors provided between one and four arguments in an article and a total of 14 separate arguments were identified by this analysis. The most frequently listed argument was that the model consisted of predictors measured at one level.

A thematic analysis of the arguments produced three themes: (a) arguments that regression and hierarchical linear models produce results that are equivalent, (b) arguments related to the composition of predictors in the model, and (c) arguments related to the variation across groups. The theme that appeared most frequently was that regression and hierarchical linear models produced results that were sufficiently similar
to treat them as identical. Some authors included language to the effect that they had constructed the model using both techniques and that the significant predictors were the same. The third theme identified by these analyses used the magnitude of variation across groups of the study as a justification for OLS regression. Perhaps reflecting the practice described by Ethington (1997), scholars calculated the proportion of total variance that can be attributed to between group differences, the ICC. The results suggest that higher education scholars have agreed that the ICC is a test that can be used for selection of analytic technique, but there is not agreement regarding the threshold for this value. A significant ICC should be interpreted to mean that at least one of the groups is significantly different from the other groups on the outcome measure. The results of this study and a general review of the literature did not identify specific guidance regarding when the magnitude of that difference may be interpreted as not meaningful. Overall, the analysis of reasons used to justify OLS regression suggested that scholars who have continued to use OLS regression believe hierarchical linear models and OLS regression are essentially equivalent analytic techniques.

Under certain circumstances, OLS regression can produce robust models. The final analysis for this sample coded the corrections, or adjustments, made that allow for the use of a single-level model. Thirty percent of articles in the *Regression Studies* sample included text to describe adjustments to the modeling process. Corrections included the use of a more conservative $p$-value, specialty software, and modifying the specified model to reduce the number of predictors or eliminate variables measured at different levels. Similar to the findings of the first part of the analysis, at least one
correction did not appear to correct the issues associated with the use of regression on multilevel data – that of dummy coding. While dummy coding can be used to control for effect of grouping, it also changes the interpretation of the intercept which was not mentioned in the discussion of results in the article.

Combined, the analyses for research question three suggested there was not clear consensus regarding the conditions under which scholars can use OLS regression on multilevel data. Scholars proffered several arguments to justify OLS regression. Some arguments, such as using predictors measured at one level, can be supported with existing statistical literature. Others, such as using a calculated ICC, were not strongly supported by the statistical literature. In terms of the hypothesis for the research question, the results are mixed. There was variation in the reasons provided by authors, but these reasons did cluster into themes. As a consequence, the hypothesis was partially supported by these analyses.

Discussion of Results

This research applied a novel design to study analytic and communications practices associated with the use of hierarchical linear modeling and multilevel data in peer-reviewed higher education journals. Each of the research questions led to a design that incorporated analyses that were qualitatively different from each other. The design, while innovative, also made it difficult to situate results in the context of existing literature. When possible, this section will ground the discussion of findings in literature. Interpretations of results are also contextualized across the various analyses conducted for each research question – similar to the process of triangulation in qualitative
methodologies (Bodgan & Biklen, 2006). This process helps address the broader question underlying this research: *What is the collective understanding of hierarchical linear modeling for higher education scholars?*

**The Literature of Hierarchical Linear Models: Expanding Our Methodological Sources**

The findings from the citation analysis supported an interpretation that the practices used by higher education scholars when reporting the results of hierarchical linear modeling represent an oversimplification of issues related to the use of the technique. The finding that 85% of articles cited at least one of the texts by Bryk and Raudenbush (1992, 2002) combined with the relative concentration of journal sources in the education/psychological journals could be interpreted to mean that higher education scholars have not explored the conceptualization of multilevel analyses in other disciplines. While some may conclude “we are on the same page” others may wonder “what are we missing?” regarding the technique.

It was not possible to infer which conclusion is most accurate from a citation analysis. Perhaps this finding is a function of how higher education researchers conceptualize the nature and focus of their research. Are questions of method and analysis merely the tools that make it possible to study the content of interest? If that were true, it is more likely that scholars would draw primarily from the literature of the field. This interpretation would be consistent with the citation analyses of this research, which identified a small set of sources on which scholars relied most heavily.
Publishing, an important activity for professional success for faculty, may also influence the citation practices of higher education scholars. Do higher education scholars select sources they expect to be familiar to editors and reviewers with the expectation that unfamiliar sources could lead to a negative outcome? A third interpretation of the results of the citation analysis is that some in the field do not possess the depth of statistical and methodological training to feel comfortable reading and citing literature based on statistical and methodological theory. As a result, scholars may rely on sources they know have been accepted as credible and cite what has appeared in other literature.

Because there is so little published in the higher education literature on citation and source use, it was necessary to look outside the literature of the field to identify and interpret other findings of the citation analysis in this research. The fields of library science and general scientific research can serve as starting points to understand the literature related to citation and source use in scholarly publications. Journal impact factors, which measure frequency with which other scholarly publications cite articles published in a particular journal, are reported by most scientific journals and are increasingly included in summaries for individual journals, including those in higher education (c.f., Research in Higher Education, Review of Higher Education). It is beyond the scope of the current research to present a lengthy discussion of the merits of citation measures and indices. However, the impact of a journal and/or a source is a concept to consider when interpreting these results.
Using the language of impact and the results of the citation analysis, Bryk and Raudenbush have the highest impact on this field’s constructed understanding of hierarchical linear modeling. Ethington (1997) had the second greatest impact. Another metric, published by Thompson-Reuters, an indexer of journal publications, is the cited half-life of a journal. The cited half-life is a measure how long after publication an article is cited in other publications. The results of this citation analysis suggest that the half-life of the identified sources is very long. The publication years of the sources cited at least five times ranged from Burstein’s 1980 article that introduced the concept of multilevel modeling to the field of education and psychological research to a text by Luke and a technical manual for the software package, HLM, both published in 2004. Excluding the Burstein article, the most frequently cited sources were published between 1992 and 2004. While this is a 13-year span of time and the sample used in this analysis includes articles going back to 2000, there has been a clear increase in the number of published articles using hierarchical linear modeling starting in 2010. It is somewhat puzzling that higher education scholars seem comfortable using a set of sources that may not represent the most current understanding of hierarchical linear models from the statistical theory.

One may argue that the half-life of articles will necessarily be shorter than those of books and observe that six of the nine most cited sources are books. Perhaps it was acceptable to identify these texts as key sources worth citing in the narrative. Alternatively, one may also argue that there are no good examples of book sources that have been published since 2004 to replace those most frequently cited by higher
education scholars. However, there are several examples of more recent texts, including Hox (2010), Gelman and Hill (2007), and Heck and Thomas (2009) that could serve as alternatives to those identified by this research. Only the last, Heck and Thomas, appears in the complete list of sources for the citation analysis (Appendix G).

The results of the *HLM Studies* sample content analysis may provide additional insight into why higher education scholars rely on the literature of Bryk and Raudenbush. Of the studies that identified which software was used to estimate the coefficients and variance components of the model, HLM software was used more frequently than other software packages available. As described in Chapter Two, this software was developed by a team that included Bryk and Raudenbush. Does the selection of a specific software package influence the literature scholars cite when writing about the technique? This interpretation seems plausible and also may help explain why higher education scholars used hierarchical linear models more frequently than what was reported by Dedrick et al. (2009). The motivating example for Bryk and Raudenbush in both their 1992 and 2002 editions is a 2-level nested model. The use of Bryk and Raudenbush may lead to practices in the application of the technique that become mutually reinforcing at multiple levels: citation, application, and study design.

Collectively, the results of the citation analysis raise several questions about the “why” of source selection that cannot be answered from this analysis. Why do higher education scholars rely on such a small set of sources? Why do these sources not represent the most current publications related to the method? Given the patterns identified by this research it seems reasonable to infer that journal submission and peer
review processes have a contributing role in the use of a short list of sources and approaches when writing about studies that use hierarchical linear models. It is possible that scholars view hierarchical linear modeling and, perhaps analytic techniques in general, in a manner that is different from their perception of content knowledge. Methodology may be perceived as a knowledge base that is static and, therefore, it would be less necessary to update methodological sources the way one might scan the literature to ensure one’s understanding of the content area reflects the most current theory and research. While this may be true for some analytic techniques that have existed for multiple decades, hierarchical linear modeling is a technique that continues to evolve so there may be relevant knowledge that should be incorporated into our understanding of the technique and how to interpret results.

The Belief that OLS Regression and Hierarchical Linear Models Are Equivalent

Implicit in the design and context for this research was a question that asked whether scholars were selecting the appropriate technique for modeling continuous outcomes on multilevel data. The general upward trend in the number of studies that used hierarchical linear modeling over the time span of articles included in this research suggested some higher education scholars are shifting their preference of analytic technique to hierarchical linear models. This shift is not universal, as the number of studies that used OLS regression did not exhibit a downward trend over the timespan of articles in this research.

It is possible that the scholars who have continued to use OLS regression were basing that decision on a few key articles, such as the comparison of OLS regression and
a multilevel model by Astin and Denson (2009), or reading studies in which scholars stated that models were constructed using both techniques and the results were equivalent. Perhaps influenced by the canon of literature that was identified in the citation analysis, scholars using empirical comparative studies are neglecting the studies that produce an empirically-based counter example. Unless scholars apply both techniques to their data, they cannot claim with complete certainty that the results are, in fact, equivalent.

The analyses of research questions two and three examined the arguments presented to justify the selected analytic technique. If OLS regression and hierarchical linear models are equivalent, even when limited to specific conditions, we might expect a pattern of justification to converge around a common set of arguments and conditions under which OLS regression is acceptable. This would suggest scholars have reached agreement regarding the conditions in which OLS regression is an acceptable alternative to hierarchical linear models for use with multilevel data.

Users of hierarchical linear models identified, correctly, that nested data or theoretical models that include predictors measures at different levels justify the use of multilevel analytics techniques, including hierarchical linear models. Some OLS persisters, correctly, argued that a model consisting of predictors measured at only one level may be analyzed using OLS regression. However, this interpretation cannot be generalized to include all studies that model continuous outcomes on multilevel data. OLS regression can be used when adjustments are made to the statistical analyses and the sample is random at both levels of the data (Heeringa, West, & Berglund, 2010). Other
OLS persisters seem to rely on anecdotal evidence rather than on statistical arguments. This is reflected in the prevalence in the frequency of uncited arguments from the Regression Studies sample content analysis.

The results suggested that the community remains divided in its beliefs about the reasons we should/should not use hierarchical linear models. Perhaps trying only to convince the reviewer, some authors have elected to include multiple and varied justifications. A skeptical reader may interpret this practice as “throwing spaghetti” in the hope that at least one argument will be acceptable to the reviewers of the journal. This seeming division of OLS regression persisters vs. HLM users may also have origins in the social nature of the field. There is a pattern of studies coming out of HERI continuing to use OLS and NSSE studies using multilevel techniques. It may not be coincidental that the 2009 article by Astin and Denson originated at HERI. The literature on academic disciplines describes the social construction of discipline based on shared values and beliefs. This pattern may be evidence of organizational influences on scholarly practices.

The results of the analysis using the Regression Studies sample illustrate issues specific to the types of studies and data used in higher education research. When Ethington published her 1997 study that described how hierarchical linear modeling could be useful to higher education researchers, her core argument was that clustered data were better analyzed using multilevel modeling. The framing of this argument may have oversimplified the comparisons between regression and hierarchical linear models. The analysis of research question three suggests that some, but not all, higher education
scholars understand that this argument may have been too simple. The following elements of the study design should be considered when selecting analytic techniques: the structure of the data set, the measurement levels of predictors one wished to include in the modeling process, and access to corrective tools.

It is important to remind the reader that more than 50% of articles in the *Regression Studies* sample made no reference of an argument to support the choice of analytic technique. It is also important to acknowledge this result was similar to that of the analysis of the *HLM Studies* sample. This result should not be interpreted to mean that scholars are not justifying their methodological choices, but that they have not reported this justification or deemed it necessary to do so. This finding suggests that there may be a need to look more carefully at the reporting practices of higher education scholars across other types of studies.

**The Effect of Journal Guidelines**

The results described in Chapter Four of this research revealed reporting patterns both within- and across-journals represented by the articles in the samples. A majority of articles omitted details about data preparation and diagnostic testing, and authors rarely addressed the assumptions of multilevel models explicitly in the text. For example, while it was a common practice to report the means and standard deviations for continuous measures, this did not provide sufficient information to determine that the assumption of normality was met for the model. The issue of missing data was reported more frequently and, when addressed, included explicit language to describe how the problem
was managed prior to analyzing the data. As discussed in Chapter Four, solutions included listwise deletion and the use of multiple imputation techniques.

In addition to the omission of details regarding data preparation and diagnostic testing, the manner in which models were reported in the narratives were not always easy to discern. The use of hierarchical linear models makes it possible to understand the complex relationship between individuals and the groups to which they belong through the specification of variance structures. To interpret these relationships, it is necessary to define the variance structures so that the reader can determine which level-1 slopes were allowed to vary across groups (McCoach, 2010). The results of this analysis showed that a majority of articles in the *HLM Studies* sample did not define the variance structure in a narrative form and frequently included it only in tables of results with no explanation.

Why did higher education scholars seem to agree on the omission of information about data preparation and present little agreement regarding how to report the coefficients and variance structures of the models? One means of explanation for these findings is to situate them in the context of the journals. Submission guidelines for the journals that served as a source of data for this research provide a possible explanation. Each of the journals included in this research lists submission guidelines for authors.

Three of the journals stated explicitly that manuscripts should conform to the style guidelines of the Publication Manual of the American Psychological Association (ACPA, 2011; Journal of Higher Education, n.d.; The Review of Higher Education, n.d.). *Research in Higher Education* did not reference a particular style manual. They did, however, include sample citation and reference formats that were consistent with the
APA formatting, so authors might interpret this to mean that the APA guidelines were appropriate. The appendix of the sixth edition of the APA Publication Manual includes tables that describe reporting recommendations regarding general content and include specific guidance for studies based on original data. A comparison of the results of this research to the APA guidelines showed that articles in *HLM Studies* sample included some, but not all, of the content recommended in the guidelines associated with data preparation.

There are several reasons why scholars may elect to omit details about model structure in submitted articles. Author guidelines for *Journal of Higher Education* state that because the journal’s readership is broad, “statistical material [should] be presented as briefly and simply as possible” (JHE, n.d.a). The difficulty of accepting this as evidence of sufficient documentation for reporting methodology lies in other statements in the APA Publication Manual. The publication manual includes a call for more detail: “…a complete description of the methods used enables the reader to evaluate the appropriateness of our methods and the reliability and validity of your results. It also permits experienced investigators to replicate the study” (APA, p. 29) and cautions that “[i]nsufficient detail leaves the reader with questions” (APA, p. 29). It is important to note, however, that the authors of the APA guidelines are seeking a balance in methodological narrative. In the same sentence, the APA cautions “too much detail burdens the reader with irrelevant information” (APA, p. 29).

Despite these cautions, the APA guidelines offer no clear support for the omission of this information from the articles in the sample. However, the publication manual
provides examples of methodological content from the Communications Board Working Group, a group formed to review reporting practices in preparation for the release of the 6th edition of the *Publication Manual for the American Psychological Association*. In the appendices of the 6th edition of the Publication Manual the authors include a description of the Journal Articles Reporting Standards (JARS). The *Review of Higher Education*, in contrast, states only that a manuscript should conform to guidelines in the APA Publication manual (Review of Higher Education, n.d). *Research in Higher Education* offers no specific instructions regarding content, nor does it require formatting of content according to the APA publication manual. Finally, the *Journal of College Student Development* makes no specific mention of required methodological content (ACPA, 2011). After reviewing the submission guidelines for each of the four journals included in this research, there is some evidence to suggest that submission guidelines inhibit the inclusion of methodological details. Failure to understand the implications of these omissions could result in misinterpretation of results. Scholars wishing to use published studies as examples of acceptable practice for conducting or reporting studies that used hierarchical linear models may be better served looking elsewhere for these examples.

As was the case with the data preparation and diagnostics, it seemed reasonable to look to the literature to explain this finding. Why would authors report information about variance minimally, or not at all? Author instructions for each of the journals did not provide insight into this finding, although given that hierarchical linear modeling is a specialized and advanced technique, detailed instructions might not be unexpected as author guidelines are designed to accommodate a breadth of research topics and designs.
The APA guidelines include a sample for reporting hierarchical linear models in table format. This format includes reporting each of the models tested (unconditional, level-1 only, full model) and included both coefficients and variance components (APA, 2010). The results of this research suggested that the more common practice in higher education is to exclude information about variance components from results in table form. A 2008 article titled, Reporting Standards for Research in Psychology: Why do we need them? What might they be? and the source for the appendices described earlier in this section, included recommendations for new data collection and meta-analyses (JARS Working Group, 2008). Secondary and non-experimental designs were not addressed in this document.

The question, then, becomes what is the appropriate level of detail and who should make that determination? The findings and variation in reporting practices represented by the articles in the HLM Studies sample suggest the current practice is to allow the author to make that determination and that there is no agreement regarding what is essential content when reporting results. For analytic approaches that have a longer history of use in higher education research, minimal methodological content may be acceptable as implicit agreement about how to use the technique may have been infused into the field in many areas and levels, including graduate program curricula and training. This would represent a stable perception of the technique by scholars. The results of this research further suggest that the field’s understanding of hierarchical linear models has not yet stabilized. The finding that scholars are presenting conflicting arguments regarding the appropriateness of OLS regression and hierarchical linear
models with complex data is evidence of this instability. The variation in table
formatting of coefficients and variance components serves as additional evidence of this
interpretation, even if reporting practices take the specific foci and research questions
into account.

The discussion of findings presented here identified three key themes of this
research. First, the “canon” of the field associated with hierarchical linear models is well
and narrowly defined given that the technique evolved across multiple disciplines and
fields simultaneously. Second, a subgroup of higher education scholars believe there is
sufficient evidence to support arguments that OLS regression techniques are equivalent to
hierarchical linear modeling. The implications of continued use of OLS regression are
discussed in the next section. Finally, the results of this research provide evidence that
submission guidelines and review practices influence the manner in which scholars report
results of studies based on hierarchical linear models and the choices made when
selecting between OLS regression and hierarchical linear model.

**Implications and Recommendations**

**Selection of Analytic Technique**

The analysis of justifications for and against the use of hierarchical linear models
on multilevel data led to the finding that there is a persistent belief that OLS regression
models are a legitimate option under a much broader set of conditions than is supported
by the statistical literature. The statistical literature supports the claim that a hierarchical
linear model is appropriate when one wishes to model a continuous outcome using a data
set with multilevel structure (Burstein, 1978; Burstein, Linn, & Capell, 1978; Gelman &
This research did not identify any sources that suggested an OLS regression model would be universally preferable to a hierarchical linear model. The results also suggested scholars preferred OLS regression to a multilevel model for two reasons. While not used as an argument to justify OLS regression, at least one scholar expressed a preference for OLS regression due to the ability to add predictors in blocks and the ease of calculating incremental increased in total variance explained. It must be acknowledged that explaining change in variance explained in a multilevel model is much less straightforward than in an OLS regression model. However, the statistical literature does not provide clear evidence that this justifies the use of OLS regression.

One of the observations from this research was that the peer-review process is not immune to error. In addition to simple citation errors such as the misstatement of authors or years for a particular source, at least one of the articles in the Regression Studies sample included a justification that was incorrect – the author claimed predictors were measured only at level-1 but included at least one variable measured at level-2. This was not identified or corrected through the peer-review process. Scholars with a preference for OLS regression should consider the following criteria when selecting between OLS regression or hierarchical linear models.

A specified model that includes predictors measured at different levels (e.g., values associated with a student and values associated with the institution they attend) should be modeled using a hierarchical linear model. Studies such as Astin and Denson (2009) that demonstrated empirically one example in which the significant results were
similar are not sufficient evidence to prove that the results will always be similar.

Proving a statement to be true cannot be established with certainty from empirical examples (Fraleigh, 1982; Mendelson, 1987). However, establishing a knowledge claim to be untrue can be proved with one counterexample (Mendelson, 1987; Simpson, 1951). This is an example of Simpson’s paradox which showed, mathematically, that inferences drawn when data are analyzed at one level can be reversed when analyzed at a different level. A model with predictors measured at different levels should be modeled using multilevel techniques.

Data sets created using sampling procedures that violate the assumption of independence (i.e., non-random at one or both levels) are best analyzed using multilevel techniques. These data sets were described as ‘nested’ by several authors in both the HLM Studies and Regression Studies samples. Articles from both samples used the ICC as a test to determine the need for multilevel techniques. Analyses of both the HLM Studies and the Regression Studies samples showed that scholars referenced a calculated value for the ICC, and may have included a citation for the formula used in the calculation. Scholars should calculate the ICC to understand the proportion of variation attributable to between group differences as it may affect the interpretation of results. However, I found no evidence in the statistical literature to indicate that a small ICC should override the arguments associated with model structure or the violation of the assumptions of independence. The results here suggest that scholars should be discouraged from using a calculated ICC as justification against the use of multilevel models with nested data.
Studies that are designed to estimate a model consisting of predictors measured at one level and that used complex random sampling procedures (i.e., the data are nested) can be analyzed using OLS regression, but ought to employ a statistical correction for this (Thomas & Heck, 2001; Heeringa, West, & Berglund, 2010). These corrections include the use of design-based approaches such as weighting procedures. Thomas and Heck, for example, posits that single level analyses “can be maintained after adjustments are made for the sample design effects” (p.521) but that this constrains the analyst to conducting an analysis using predictors measured at either level-1 or level-2 but not both. Weights are incorporated into the modeling process to adjust for the distribution of respondents on key demographic characteristics. As summarized in the review of the literature, the use of an adjusted sample size is another strategy that will yield a “more accurate standard error” (Thomas & Heck, 2001, p. 533). Because the degrees of freedom are reduced in nested data, the number of predictors included in the model should also be reduced. In accordance with the recommendations of Thomas and Heck, the use of more conservative p-values should be considered the least desirable correction.

**Reporting Results**

The results of this research showed there was wide variation in the reporting practices associated with hierarchical linear models. Comparisons were made to a set of proposed standards (termed desiderata) in Chapter Four (McCoach, 2010). These standards were not published until recently so it was not possible to argue higher education scholars have not aligned their practices to a recommended best practice. An earlier section of this chapter explained some of this variation using the context of journal
submission guidelines and found that submission guidelines may influence how scholars report methodology and analyses. These explanations did not, however, resolve the practical implications of this research. Hierarchical linear modeling is a still emerging technique of the field and the absence of a generally accepted reporting approach makes it difficult for untrained readers to critique the interpretation of results. Until the technique is better established in the field, it is recommended when possible, that scholars adopt a standard approach to reporting the results of studies using a hierarchical linear model.

There are several means of creating this standard. It could be based on the guidelines proposed by McCoach (2010) and used in the analysis of this research. Higher education scholars, for example, may conclude that the field is best served by excluding information about data preparation. Assuming the results of this study are reflective of broader practices of the field, including minimal information about data preparation would be consistent with current practices but in contradiction to the standards proposed by McCoach. The identification of software and estimation methods and algorithms did not occur regularly in HLM Studies sample. Scholars may elect to continue to omit this information. However, in doing so, the value of published studies using hierarchical linear models is diminished as a tool for replication. A long term consequence of such a choice may be that it becomes more difficult to explain study results that contradict results of prior research, and they become less valuable to scholars wishing to use these studies to learn how to use hierarchical linear models as a research tool.
Professional associations could engage in a collaboration to develop a reporting model for hierarchical linear models. If there is no interest in a collaborative approach, which seems possible given the submission guidelines for the journals represented here, then it is recommended that the editorial board for *Research in Higher Education* develop these guidelines. *Research in Higher Education* is the logical source for this work for several reasons. First, *Research in Higher Education*, a publication of the Association for Institutional Research, has a focus on method. Second, this is the journal that published a greatest proportion of articles included in the *HLM Studies* sample. Editors and peer reviewers at *Research in Higher Education* may have deeper understanding of the technique than those of the other journals, which have included explicit statements in their publication guidelines that discourage inclusion of methodological content in favor of topical content. Those involved in the development of reporting guidelines could structure recommendations in a manner similar to the work of the JARS group described previously or published as supplemental guidelines included to the APA manuals (American Psychological Association, 2010; JARS Group, 2008). These guidelines should be disseminated publicly in accessible locations such as journal publications and association websites.

The benefits of such guidelines are multiple. First, the process may create a forum to discuss and document the ways in which hierarchical linear models are qualitatively different from OLS regression. This may reduce the incidence of methodologies that were not applied consistently by scholars such as the use of the ICC as a means test for the selection of OLS regression or a multilevel model. For those
unfamiliar with hierarchical linear models, standardized reporting makes journal articles a resource for learning about hierarchical linear modeling as a tool of research and reinforces a best practices approach to scholarly communication.

Finally, the use of a standardized reporting approach may help address what appears to be a limitation to the peer-review process. This research identified some factual errors that were not corrected through the peer-review process. In one example, the author made a statement about the specified model that was inaccurate. This error was not identified by an existing peer-review process. In the case of this example, for this author, the error about the specified model may have affected the choice of OLS regression or hierarchical linear model. This type of error is qualitatively different than a misidentified publication year and could affect our understanding of the focus of the article.

The development of standard reporting guidelines may not be without controversy. Scholars may believe that their work is compromised if reporting standards are encouraged. It is also possible that some of the decisions to include or exclude content are driven by external pressures, such as page limits, described in submission guidelines. The peer-review process may also need to be reviewed to ensure article is reviewed by someone with experience in the technique.

**Strengthening the Tools of Our Research**

One of the challenges to applied fields identified in Chapter One was that the tools of our research have become increasingly complex. The value of these tools is clear: there is much we know about higher education that would not be possible without
advanced techniques such as hierarchical linear models. The challenge to the field, however, is to manage the need for content acquisition against deeper understanding of analytic tools. The finding that higher education scholars are not searching widely in the literature when educating themselves on a technique suggests if we want higher education scholars to learn more about these tools, the information must be contained within the literature of the field. Resources published in related fields such as economics, political science, or public health were not likely to be used by higher education scholars.

The question becomes how do we create spaces in the higher education community to better educate ourselves about advanced tools of research? One strategy would be to publish more articles about methodology in the leading journals. In 2006, Pascarella argued that the rate of knowledge production in higher education made it impractical to produce literature reviews similar to the ones in How College Affects Students (Pascarella & Terenzini, 1991, 2005). Pascarella argued that there was a need to publish literature reviews for the field to manage the amount of content knowledge that was created each year and that journals were the appropriate location for these reviews as the time to publication for books made a review in book form out of date by the time it was published. Unfortunately, it appears that these reviews have not been conducted. Scans of the literature in the journals represented by this research did not show any meaningful increase in the number of articles that would be classified as a literature review. The predominant article type continues to be empirical research.

This is somewhat unfortunate because this research demonstrates a need for broad reviews of method that scholars could utilize as a foundation for empirical studies. For
example, this research showed that the field may benefit from a publication that synthesized the issues associated with selection of OLS regression or hierarchical linear modeling. This may eliminate the differing perceptions of the technique that emerge only when conducting an analysis such as that of this research. It would also make it possible to benchmark the field’s perception of the technique against other disciplines and the statistical literature. The quality of some articles in *HLM Studies* sample showed that higher education has active scholars possessing the necessary knowledge to create these documents. They could help translate the language of statistics so that higher education scholars apply research tools in meaningful and accurate ways within the field.

To help reinforce the importance of methodological quality, publication guidelines should be revised so to allow scholars to include more methodological content in articles. Page limits are artifacts of print publications. With the vast majority of articles being accessed online, page limits are no longer as critical as in the past. One method for managing a revised page limit is to encourage scholars to provide detailed methods as an appendix that is available online only or by request for those without online access. This appendix would serve as evidence of method for those wishing to learn more about how to conduct such analyses or replicate existing studies. They would also be useful to beginning scholars wanting or needing to learn how to effectively communicate their work.

Publishing serves multiple purposes in a community of scholars. It serves as an important means of communicating new knowledge. Publishing in peer-reviewed journals is accepted as evidence of scholarly productivity in faculty reward structures for
promotion and tenure. Peer-review is an embedded form of implicit validation of the quality and relevance of scholarship. The process of peer-review, however, is not without critics. There are several critiques of peer-review in the broader literature including general science (Jennings, 2006). These critiques recommend several lenses that might be used to understand how peer-review affects scholarly communication. The findings of this research are evidence that the publication and peer-review process may present structural barriers to scholarly communication in higher education.

The findings of this research also suggest applied fields such as higher education should consider requiring at least one reviewer of a submitted article have substantial experience with analytic technique of the article. As part of contextualizing the findings of this study, the author examined submission guidelines for each of the journals represented in the samples to identify explicit instructions that helped to explain the results. This was discussed in a previous section of this chapter. As part of that review, efforts were made to locate information regarding the review process for each journal. The information available was limited. Lists of editorial board members were available, but little that described specifically how reviewers were assigned, their credentials, or roles in their work. Expecting a reviewer to have deep knowledge of the topic seems reasonable. It also seems reasonable that not all scholars will have understanding of analytic technique. This may help explain the variation in how authors have justified hierarchical linear models.

The challenge to implementing such a recommendation is that it may require a change in espoused and enacted values reflected in the submission and review process for
by the editorial boards of the journals included in this study. Publishing is a high stakes process for authors and journals alike. It may be difficult to enact the changes proposed here without a thorough review of the effect such changes would have on the journal and authors.

The recommendations of this study identified several opportunities to strengthen the quality of the scholarship of our field. While many of these recommendations are directed at hierarchical linear models, they also have application to other advanced analytic tools and apply to both qualitative and quantitative methods.

**Limitations**

The results and interpretation of this research are moderated by several limitations. Five limitations have been identified and are discussed here.

First, the samples used for the analyses did not include all studies by higher education scholars that use hierarchical linear modeling. Studies have been published in other journals (e.g., *Journal of Diversity in Higher Education*), which may yield different results. However, citation analysis of the field indicates that researchers rely heavily on the four major journals included in this dissertation (Budd & Magnuson, 2010), which suggests that unless journal submission guidelines are substantially different from the those included in this research, we would expect results to be similar to those presented here.

Second, while it seemed reasonable to apply the coding framework developed by Dedrick and colleagues (n.d.) to a sample of articles with a higher education focus, it was possible that the codebook would not translate to the sample used in this research. As
discussed in Chapter Four, the codebook for this analysis was modified during the coding process. The other two research questions required similar modifications to coding processes as articles were analyzed for each research question. This was needed to provide a context for use in interpreting and making meaning from the pre-defined codes of the study. Qualitative methodologists might argue that this is consistent with a constant comparative method--revising and modifying codes as the analysis was conducted (Glesne, 1999). However, this study was designed to utilize deductive coding procedures, not inductive. While I believe these modifications increased the overall quality of the study and provided stronger evidence of the findings, interpretations, and implications, others may find that the revision of research protocols for content analytic work to be a limitation.

A third limitation of this research is that it was conducted by a single reader. Content analysis is typically conducted using multiple readers and uses statistical measures of reader agreement to establish the reliability of coding procedures and validity of findings (Krippendorff, 2004). A typical strategy for minimizing the potential bias includes calculating reliability measures for coding used in the study. This is achieved through the process of training multiple readers to code a subsample of documents in the entire sample and calculate statistics to measure the level of agreement in coding choices. Krippendorff recommended Krippendorff’s alpha (2004). While not unheard of, it is still a general practice that dissertation research is conducted independently by a doctoral candidate. The modifications described in Chapter Four and the approach to content analysis used in this study incorporates a level of interpretation
that could bias results. The focus of the analysis in this study, hierarchical linear models, required an advanced knowledge of statistical techniques. The expertise needed and the fact that each research question required a different approach to coding and interpreting article content made it difficult to identify suitable coders for this research. However, efforts have been made by the author to design coding strategies and procedures that minimize interpretation and bias of results. For example, each article was read and coded at least twice on a separate coding sheet with at least one week between readings. Final code determinations were made by comparing the coding from each reading. Early in the coding process, there were frequent differences across codes for the same article. As I became more familiar with the coding process, coding became more accurate across the repeated reading of an article and there were few or no differences. In addition, findings were triangulated across the analyses for each research question. It is possible, however, that a different reader would produce different coding.

The Regression Studies sample for this research was constructed using criteria that were intentionally restrictive. The sample excluded articles based on data with fewer than 10 level-2 groups and/or those that modeled non-continuous outcome measures. Not surprising, then, arguments based on the number of groups was largely absent from the justification and corrections identified in research question three. In retrospect, I would not have excluded articles based on data sets with fewer than 10 level-2 groups. While this would have added only three articles to the Regression Studies sample, it may have shifted the patterns of justification given by authors when arguing in favor of OLS
regression for their analytic approach. Such an extension may be appropriate in future research.

The scope of this research does not include all scholarly articles that used other types of multilevel analytic techniques. Although scholars sometimes used the term, \textit{HLM}, to represent the entirety of multilevel modeling techniques, in this research only those studies that constructed a multilevel model on a continuous dependent measure - the definition of a hierarchical linear model - were included in this research. The reasons for this decision were detailed in prior chapters. However, the consequence of this design choice is that these results are not generalizable to other types of multilevel analyses such as those that model binary or categorical outcome measures. It is also possible that the results of the citation analysis would be different if that analysis included all multilevel analyses.

Finally, it is possible that the years represented by the articles included in the \textit{HLM Studies} and \textit{Regression Studies} samples are indicative of the period in the evolution of a field that is characterized by disagreement about a topic or tool of research. It may be that including studies published since mid-2012 would produce different results regarding the level of agreement within the field regarding reporting practices or closer alignment with the standards proposed by McCoach (2010).

\textbf{Recommendations for Future Research}

The analyses in this research yielded findings that present several opportunities for further research. This is due, in part, to the exploratory nature of the study and that there is little published literature against which to compare the results. The results of this
research describe the “what” of this topic; future research should investigate the “why” and “how” of the topic. What reasons, for example, would higher education scholars give for their continued use of OLS regression models? What factors influence the process of selecting sources for the purpose of citation? How do higher education scholars learn new analytic tools after completing their formal training in a doctoral program? What are the central characteristics of methodological and research training in higher education doctoral programs? How can we infuse the statistical literature into our field, and is this necessary to strengthen the quality of our scholarship?

Future research should continue to develop a knowledge base regarding how and what scholars report in journal articles. In what ways does the journal selected for submission affect how methodology is communicated? Do journals affect citation patterns? What role do submission and style guidelines have on the general reporting practices of scholars? How do guidelines and practices differ by journal and what effect does that have on our communal understanding of topics both methodological and that which would be considered traditional content?

The present study could be replicated using a different article samples or analytic techniques. Would a study of articles using structural equation modeling, for example, produce similar variation in reporting practices? Or does the “age” of the technique in the field lead to the stabilization of reporting practices as was suggested earlier in this chapter? This line of research would provide evidence to determine if the results of this study are unique to hierarchical linear models or representative of higher education scholarship and its use of statistical techniques in general. The results of research
question three, for example, showed that not all justifications and corrections were explicitly supported by sources and the results of research question one showed that scholars, in general, included few sources related to method. How do scholars select their sources? Is it an artifact of their academic and professional preparation?

Chapter Summary

The purpose of this research was to examine the reporting practices of higher education scholars in studies of complex survey data and hierarchical linear modeling. By examining two types of studies – hierarchical linear modeling and regression models – this research provides a foundational understanding of how scholars are reporting the results of these studies and identified several opportunities for higher education scholars to strengthen the methodological quality of their work that applied to hierarchical linear models and to generally accepted research methods.
APPENDIX A

HLM STUDIES SAMPLE SOURCES
Note: The reference list consists of the articles included in the analysis for research questions one and two.


APPENDIX B

HLM STUDIES CITATION ANALYSIS CODE BOOK
HLM CITATION ANALYSIS CODE BOOK

1. Sample Article Characteristics
These codes are assigned to each article in the sample. Note: Not all coding is reported in the results.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description (if needed)</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Article Code</td>
<td>Use naming convention from Atlas.ti records</td>
</tr>
<tr>
<td>1.1</td>
<td>Publication Year</td>
<td>Select a year 2000-2012, inclusive</td>
</tr>
<tr>
<td>1.3</td>
<td>Author(s)</td>
<td>List all author(s) for the article</td>
</tr>
<tr>
<td>1.4</td>
<td>Total # of Sources in the Reference List</td>
<td>Integer</td>
</tr>
<tr>
<td>1.5</td>
<td>Total # of Sources in Methods related sections</td>
<td>Integer</td>
</tr>
<tr>
<td>1.6</td>
<td>Total # of Sources related to hierarchical linear modeling</td>
<td>Integer</td>
</tr>
<tr>
<td>1.7</td>
<td>Citation List</td>
<td>List all citation(s) related to code F (hierarchical linear modeling citations)</td>
</tr>
</tbody>
</table>

2. Citation Source Characteristics
Note: Create a master list of all citations identified in Code “G” of the Sample Article Characteristics. Coding is assigned to each citation.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description (if needed)</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Source for Citation</td>
<td>Use APA formatting</td>
</tr>
</tbody>
</table>
2.2 Citation Purpose

<table>
<thead>
<tr>
<th>Inferred intent/purpose of citation in text</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (A) Analysis</td>
</tr>
<tr>
<td>b. (ICC) Intraclass Correlation</td>
</tr>
<tr>
<td>c. (J) Justification</td>
</tr>
<tr>
<td>d. (U) Unit-of-Analysis</td>
</tr>
<tr>
<td>Problem</td>
</tr>
</tbody>
</table>

2.2 Source Type

| a. (B) Book |
| b. (BC) Book Chapter |
| c. (M) Technical Manual |
| d. (J) Journal Article |
| e. (R) Report |
| f. (C) Conference Paper |
| g. (W) Webpage |
| h. (O) Other |

3. Journal Source Analysis

Note: This coding is applied to those sources with the code “J”, journal article assigned to 2.2 Source Type.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description if needed</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Name of Source Journal</td>
<td></td>
<td>Open ended response</td>
</tr>
</tbody>
</table>

A. Journal Discipline

Assign journal to a discipline/field based on terms in journal title

Open ended response

Final List:

- (Educ) Education
- (Psych) Psychology
- (Soc) Sociology
- (Stats) Statistics
- (Ed Psych) Educational Psychology
- (Hi Ed) Higher Education
- (Public Health) Pub Health
- (Ed Stats) Educational Statistics
- (Management) Management
- (Res Methods) Research Methods
- (Economics) Economics
- (Medical) Medical

HLM Studies Citation Analysis Coding Form

Reviewer

<table>
<thead>
<tr>
<th>Code Label</th>
<th>Response Options</th>
<th>Page #</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Citation List

Note: List all citations related to methodology or in the methods subheading of the article

<table>
<thead>
<tr>
<th>Citation (in apa)</th>
<th>Original Source Confirmed in Endnote</th>
<th>Purpose (use A, ICC, J, U, other codes) (list all that apply)</th>
<th>For A, ICC, J, U, source type (B, BC, M, JA, R, C, W, O)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C

HLM STUDIES CONTENT ANALYSIS CODE BOOK
HLM Content Analysis Code Book  
Adapted from Dedrick et al. (n.d.)

Note: These codes are assigned to each article in the sample. Not all coding is reported in the results. Coding is recorded in Atlas.ti 7.0.85 and Coding Form, and transferred to an Excel 2010 spreadsheet.

### 1. Sample Article Characteristics

<table>
<thead>
<tr>
<th>Code</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Article Code</td>
</tr>
<tr>
<td>1.3</td>
<td>What is the study type?</td>
</tr>
<tr>
<td></td>
<td>a. 2-level nested model (individuals nested in a group/context)</td>
</tr>
<tr>
<td></td>
<td>b. 2-level growth curves</td>
</tr>
<tr>
<td></td>
<td>c. 3-level nested model</td>
</tr>
<tr>
<td></td>
<td>d. 3-level growth curves</td>
</tr>
<tr>
<td></td>
<td>e. repeated measures</td>
</tr>
<tr>
<td></td>
<td>f. other, describe</td>
</tr>
<tr>
<td>1.4</td>
<td>Do the authors explicitly justify the selection of multilevel modeling (mlm)?</td>
</tr>
<tr>
<td>1.4.a</td>
<td>If 1.4 = Y what arguments are given?</td>
</tr>
<tr>
<td></td>
<td>0 = No</td>
</tr>
<tr>
<td></td>
<td>1 = Yes</td>
</tr>
</tbody>
</table>

### 2. Data Characteristics

<table>
<thead>
<tr>
<th>Code</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Source of Data</td>
</tr>
<tr>
<td>2.1.a</td>
<td>Is this a secondary data analysis (i.e., data collected for a different purpose than this study/research question)?</td>
</tr>
<tr>
<td></td>
<td>0 = No</td>
</tr>
<tr>
<td></td>
<td>1 = Yes</td>
</tr>
<tr>
<td>2.2</td>
<td>Sampling Strategy</td>
</tr>
<tr>
<td>2.3.a</td>
<td>How were level-1 unit reported (select all that apply)?</td>
</tr>
<tr>
<td></td>
<td>0 = Not reported</td>
</tr>
<tr>
<td></td>
<td>1= total units</td>
</tr>
</tbody>
</table>
2.3.b How were level-2 units reported? | 2 = individual group units  
| 0 = not reported  
| 1 = total level-2 units

### 3. Data Preparation/Diagnostics

<table>
<thead>
<tr>
<th>Code</th>
<th>Response Options</th>
</tr>
</thead>
</table>
| 3.2 | Did the authors discuss outliers of variables?  
0 = No  
1 = Yes, non specific  
b. predictors/independent measures only  
c. outcome measures only  
d. both predictor and outcome measures |
| 3.3 | Did the authors discuss reliability of scale variables?  
a. no  
b. predictors/independent measures only  
c. outcome measures only  
d. both predictor and outcome measures |
| 3.4 | Did the authors discuss power?  
a. no  
b. general guidelines considered  
c. power analysis conducted  
d. other |
| 3.5 | Did the authors discuss missing data?  
0 = No  
b. yes, at both levels  
c. yes, at level-1  
d. yes, at level-2 |
| 3.5.a | If data is missing at level-1, how was it handled?  
NA = Does not apply  
0 = Unknown  
c. Other – open response |
| 3.5.b | If data was missing at level-2 how was it handled?  
NA = Does not apply  
0 = Unknown  
c. Other – open response |

### 4. Testing Model Assumptions

<table>
<thead>
<tr>
<th>Code</th>
<th>Response Options</th>
</tr>
</thead>
</table>
| 4.0 | Were model assumptions discussed  
0 = No  
1 = Yes |
| 4.0.a | If 4.0 = “yes” what aspects were discussed  
Open-ended |

### 5. Model Construction and Reporting

<table>
<thead>
<tr>
<th>Code</th>
<th>Response Options</th>
</tr>
</thead>
</table>
| 5.2 | Was an unconditional/baseline model reported?  
a. no  
b. yes  
U = unable to determine |
<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3 Was the ICC reported?</td>
<td>a. no</td>
</tr>
<tr>
<td></td>
<td>b. yes</td>
</tr>
<tr>
<td>5.2.a Was a level-1 model reported?</td>
<td>0 = No</td>
</tr>
<tr>
<td></td>
<td>1 = Yes</td>
</tr>
<tr>
<td></td>
<td>U = Unable to determine</td>
</tr>
<tr>
<td>5.2.b Was the intercept modeled using level-2 predictors?</td>
<td>0 = No</td>
</tr>
<tr>
<td></td>
<td>1 = Yes</td>
</tr>
<tr>
<td></td>
<td>U = Unable to determine</td>
</tr>
<tr>
<td>5.2.c Was the intercept allowed to vary randomly?</td>
<td>0 = No</td>
</tr>
<tr>
<td></td>
<td>1 = Yes</td>
</tr>
<tr>
<td></td>
<td>U = Unable to determine</td>
</tr>
<tr>
<td>5.2.d Were any level-1 slopes allowed to vary randomly?</td>
<td>0 = No</td>
</tr>
<tr>
<td></td>
<td>1 = Yes</td>
</tr>
<tr>
<td></td>
<td>U = Unable to determine</td>
</tr>
<tr>
<td>5.4 How were predictors selected? (select all that apply)</td>
<td>a. based on <em>a priori</em> considerations (prior literature, theoretical arguments)</td>
</tr>
<tr>
<td></td>
<td>b. significance testing for individual predictors</td>
</tr>
<tr>
<td></td>
<td>c. effect sizes for individual predictors</td>
</tr>
<tr>
<td></td>
<td>d. fit statistics (e.g., AIC or BIC)</td>
</tr>
<tr>
<td></td>
<td>e. other (describe)</td>
</tr>
<tr>
<td></td>
<td>f. not discussed/unable to determine</td>
</tr>
<tr>
<td>5.6 Were interactions included in the models? (select all that apply)</td>
<td>a. no</td>
</tr>
<tr>
<td></td>
<td>b. level-1</td>
</tr>
<tr>
<td></td>
<td>c. level-2</td>
</tr>
<tr>
<td></td>
<td>d. cross-level interactions</td>
</tr>
<tr>
<td></td>
<td>e. unable to determine</td>
</tr>
<tr>
<td>5.8 Was there centering of level-1 variables?</td>
<td>a. no</td>
</tr>
<tr>
<td></td>
<td>b. grand mean</td>
</tr>
<tr>
<td></td>
<td>c. group mean</td>
</tr>
<tr>
<td></td>
<td>d. other (describe)</td>
</tr>
<tr>
<td></td>
<td>e. unable to determine</td>
</tr>
<tr>
<td>5.10 How were fixed effects (coefficients) reported? (select all that apply)</td>
<td>a. series of regression equations</td>
</tr>
<tr>
<td></td>
<td>b. single mixed model equation</td>
</tr>
<tr>
<td></td>
<td>c. list of estimated effects</td>
</tr>
<tr>
<td></td>
<td>d. verbal description</td>
</tr>
<tr>
<td></td>
<td>e. not reported</td>
</tr>
<tr>
<td>5.10a How was the theoretical model reported?</td>
<td>a. series of regression equations</td>
</tr>
<tr>
<td></td>
<td>b. single mixed model equation</td>
</tr>
<tr>
<td></td>
<td>c. list of predictors</td>
</tr>
<tr>
<td></td>
<td>d. verbal (paragraph) description</td>
</tr>
<tr>
<td></td>
<td>e. not reported</td>
</tr>
<tr>
<td>5.11 How were variance structures communicated? (select all that apply)</td>
<td>a. not mentioned</td>
</tr>
<tr>
<td></td>
<td>b. equation representation</td>
</tr>
<tr>
<td></td>
<td>c. list of estimated variance parameters</td>
</tr>
<tr>
<td></td>
<td>d. verbal description</td>
</tr>
</tbody>
</table>
### 6. Estimation and Testing

<table>
<thead>
<tr>
<th>Code</th>
<th>Response Options</th>
</tr>
</thead>
</table>
| 6.1 What software was used for the analysis? | a. not stated  
b. package only (list)  
c. package and version/year (list) |
| 6.2 What estimation method was used?      | 0 = Not stated – unknown  
1 = Not stated – implicit from software package (use only when software is explicitly identified)  
2 = Stated (describe) |
| 6.3 What estimation algorithm was used?   | 0 = Not stated – unknown  
1 = Not stated – implicit from software package (use only when software is explicitly identified)  
3 = Stated (describe) |
| 6.4 Did the authors discuss the issue of convergence? | 0 = No  
1 = Yes |
APPENDIX D

REGRESSION STUDIES SAMPLE SOURCES


APPENDIX E

REGRESSION STUDIES CONTENT ANALYSIS CODE BOOK
REGRESSION CONTENT ANALYSIS CODE BOOK

Note: These codes are assigned to each article in the sample. Note: Not all coding is reported in the results. Coding is recorded in Atlas.ti 7.0.85 and the checklist. Codes transferred to an Excel 2010 spreadsheet for analysis.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description (if needed)</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Article Code</td>
<td>Use naming convention from Atlas.ti records</td>
<td>e.g. JHE_82.2_Bozeman</td>
</tr>
<tr>
<td>3. Publication Year</td>
<td>Identify year of publication</td>
<td>Select a year 2000-2012, inclusive</td>
</tr>
<tr>
<td>2.a Journal Source</td>
<td></td>
<td>e.  <em>Journal of College Student Development</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>f.  <em>Journal of Higher Education</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>g.  <em>Research in Higher Education</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>h.  <em>Review of Higher Education</em></td>
</tr>
<tr>
<td>4. Author(s)</td>
<td>List all author(s) for the article</td>
<td>Record using APA format (e.g., Dugan, P. D.)</td>
</tr>
<tr>
<td>5. Source of Data</td>
<td>Identify the source of data used in the study</td>
<td>Open-ended response. For secondary data analysis, identify source and year if possible (e.g., BPS:96/98).</td>
</tr>
<tr>
<td>6. Explicit Multilevel Data?</td>
<td>Do the authors acknowledge explicitly the data has a multilevel (nested) structure?</td>
<td>1. Yes 0. No</td>
</tr>
<tr>
<td>5.b Level-2 Groups Reported</td>
<td>Do the authors report number of level-2 units (groups?)</td>
<td>1. Yes 0. No</td>
</tr>
<tr>
<td>5.a Explicit Multilevel Model</td>
<td>Do the authors acknowledge the model has predictors at both levels?</td>
<td>1. Yes 0. No</td>
</tr>
<tr>
<td>7. Group Level predictors?</td>
<td>Does the proposed model include level-2(group) predictors?</td>
<td>1. Y 0. N</td>
</tr>
<tr>
<td>6.a If 6=’Y’ Predictor list</td>
<td>List the predictors</td>
<td>Open-ended response</td>
</tr>
<tr>
<td>6.b If 6=’Y’ Continuous</td>
<td>Are any level-2 predictors identified in Q6.a continous (i.e., not categorical to represent institutional type)?</td>
<td>1. Y 0. N</td>
</tr>
<tr>
<td>8. Model Type</td>
<td>Term/phrase author(s) used to describe type of modeling</td>
<td>Open-ended response</td>
</tr>
<tr>
<td>9. Analysis Rationale?</td>
<td>Did the authors present a rationale for analytic technique</td>
<td>1. Y 0. N</td>
</tr>
<tr>
<td>8.a Regression Rationale</td>
<td>Did the rationale include justification of regression?</td>
<td>1. Y 0. N</td>
</tr>
<tr>
<td>10. If 8=Yes Justification Reasons</td>
<td>list reason(s) presented in text</td>
<td>Open-ended response. Final List (3/7/2013)</td>
</tr>
<tr>
<td>9.a Justification</td>
<td>How many explicit arguments did</td>
<td>Integer value (use 0 for those that</td>
</tr>
<tr>
<td>Count</td>
<td>the authors provide?</td>
<td>did not justify</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>11. Correction?</td>
<td>Did the authors apply a statistical correction that they believed addressed the issues related to applying a single level analysis on multilevel data?</td>
<td>1. Yes 0. No</td>
</tr>
<tr>
<td>12. If 10=Y Correction</td>
<td>If 10 = ‘Y’ describe the correction</td>
<td>Open-ended response Final List (3/8/2013)</td>
</tr>
<tr>
<td>11.a Correction Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Weights</td>
<td>Did the authors mention using weights in the analysis?</td>
<td>1. Yes 0. No</td>
</tr>
<tr>
<td>12.a If 12=Y, Correction weight?</td>
<td>Was weighting mentioned as a correction to the regression model?</td>
<td>1. Yes 0. No</td>
</tr>
<tr>
<td>14. Considered HLM?</td>
<td></td>
<td>1. Yes, tested 0. No mention</td>
</tr>
<tr>
<td>13.a If 13=Y modeled in HLM</td>
<td></td>
<td>1. Y 0. N</td>
</tr>
</tbody>
</table>

Created 8/20/2012
Rev: 11/21/2012 – differentiated explicit and implicit multilevel structure reporting (codes: 5, 5.a, 5.b)
Rev: 1/15/2013 – added code for predictor list and HLM and weighting ( 6.a, 6.b, 12, 13) Reworded code 6
Rev: 2/8/2013 – added code 12.a, 8.a
Rev: 3/8/2013 – added 9.a, 11.a, divided 13 into 2 parts
<table>
<thead>
<tr>
<th>Code Label</th>
<th>Code Name</th>
<th>Page #</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Article Code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Publication Year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.a</td>
<td>Journal Source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Author(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Source of Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Explicit Multilevel Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.b</td>
<td>Level-2 Groups Reported?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.a</td>
<td>Explicit Multilevel Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Group Level Predictors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.a</td>
<td>If 6=Y Predictor List</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.b</td>
<td>If 6=Y Continuous?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Model Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Analysis Rationale?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.a</td>
<td>Regression Rationale?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>If 8.a=Y Justification Arguments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.a</td>
<td>If 8.a=Y Justification Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Correction?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>If 10=Y Correction Detail</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.a</td>
<td>If 10=Y Correction Count</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Weights?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.a</td>
<td>If 12=Y Correction Weight?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Considered/Tested HLM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.a</td>
<td>If 13=Y, tested?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX F

SAMPLE SCREENING PROCEDURES
Two data sets were required to answer the research questions. The data were extracted from empirical studies published between January 1, 2000 and June 30, 2012 in the following journals: *Journal of College Student Development*, *The Journal of Higher Education*, *Research in Higher Education*, and *The Review of Higher Education*. These journals were selected because they are the most frequently cited sources in the field of higher education research with a high rate of rejection (Tight, 2008).

Because it was impossible to use on-line search functions to identify precisely the list of studies to include in this research, it was necessary to identify a pool of publications as a first step for this research. A list of eligible studies was constructed by conducting an all text search for each for each of the journals using the terms “hierarchical,” “multi-level,” “multilevel,” and “HLM”. It was assumed the probability of an article using hierarchical linear models but did not include at least one of these terms was close to zero. It seemed reasonable, therefore, that the reduced set would likely contain all studies published in these journals that used hierarchical linear models. The first stage of screening yielded 466 articles for further review. The distribution of articles is summarized in Table 21.
Table 21. Results of HLM Studies Sample Construction Screening

<table>
<thead>
<tr>
<th>Journal Title</th>
<th>N&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review of Higher Education</td>
<td>88</td>
</tr>
<tr>
<td>Journal of College Student Development</td>
<td>105&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Journal of Higher Education</td>
<td>96</td>
</tr>
<tr>
<td>Research in Higher Education</td>
<td>168</td>
</tr>
</tbody>
</table>

<sup>a</sup>N = 457.

**Construction of the HLM Studies Sample**

The initial focus of this research was only on studies that used hierarchical linear modeling. The development of the data set for these analyses consisted of two stages and required a scan of all potential articles. The first was to identify those articles that used hierarchical linear models. These articles were included in the analysis for research questions one and two. This screening was conducted using Atlas.ti 7.0. All potential articles were loaded into the software and text was searched and coded to identify all sentences that included at least one of the search terms. The coded text was reviewed in the context of paragraphs and placement in the structure of the article to determine if the article was an empirical study that used hierarchical linear models. This produced a total of 60 articles from the four journals. Table 3, included in Chapter Four of this research summarized the results of the screening managed using Atlas.ti and the distribution of *HLM Studies* sample articles across the four journals included in this research.

As part of the screening process, articles that were excluded from the final data set were coded with the reason for exclusion. Examples of these codes include ‘literature

---

<sup>1</sup> The online index for Journal of College Student Development includes only those articles published after January 1, 2003. A scan of all empirical studies published in print versions of the serial was conducted and identified one study for inclusion in the *HLM Studies* sample and four studies for inclusion in the *Regression Studies* sample.
review’, ‘qualitative study’, ‘other multilevel model’, ‘noncontinuous dependent measure.’ This coding was recorded on an Excel spreadsheet as justification for excluding an article from the *HLM Studies* sample data set. This proved useful when the *Regression Studies* sample was constructed.

**Construction of the Regression Studies Sample**

Regression modeling has been applied to data with both simple and complex structure. Because the focus of research question three was on studies that modeled outcomes on data sets with complex structures, the data set for research question three was constructed to include those studies that may have been deemed ‘HLM eligible.’ A set of inclusion criteria were developed to identify those studies that were most likely candidates for multilevel modeling. The following criteria were used to identify articles to include in the *Regression Studies* sample:

1. The study included data from multiple groups;
2. The number of level-2 groups was greater than 11;
3. The outcome (dependent) variable was continuous; and
4. The analysis included some form of regression modeling.

Starting with the list of articles that were excluded from the *HLM Studies* sample, a spreadsheet was created that listed each article that had been coded as ‘other quantitative method.’ This created the initial pool of articles subject to additional screening and review. Using the automated coding procedures in Atlas.ti, all articles included in the original pool were coded using the term ‘regression.’ It was assumed that an article that did not include the term “regression” was not likely to use regression in the
analysis. Each article that included the term “regression” was then reviewed manually to make a final determination of the article’s status for this research (inclusion/exclusion). Articles that were excluded were coded with the reason for exclusion. The codes consisted of a descriptor that was the negative of one of the inclusion criteria listed above. Once a study was determined to be ineligible based on one criterion, no additional screening was conducted and the article was excluded from further analysis. A total of 50 articles were included in the Regression Studies sample. Table 4, reported in Chapter Four, summarized the final distribution of articles included in the Regression Studies sample used for research question three.
APPENDIX G

HLM STUDIES SAMPLE CITATION ANALYSIS REFERENCE LIST
Note: References reflect original formatting in *HLM Studies* sample.


Rowe, K. J. (1999). *Multilevel structural equation modeling with MLn/MLwiN & LISREL 8.30: An integrated course*. Check ??Marsh 73.3??.


APPENDIX H

FREQUENCY COUNT FOR HLM STUDIES SAMPLE SOURCES CITED
<table>
<thead>
<tr>
<th>Source</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raudenbush &amp; Bryk (2002)</td>
<td>34</td>
</tr>
<tr>
<td>Ethington (1997)</td>
<td>20</td>
</tr>
<tr>
<td>Bryk &amp; Raudenbush (1992)</td>
<td>17</td>
</tr>
<tr>
<td>Heck &amp; Thomas (2000)</td>
<td>9</td>
</tr>
<tr>
<td>Kreft &amp; de Leeuw (1998)</td>
<td>9</td>
</tr>
<tr>
<td>Thomas &amp; Heck (2001)</td>
<td>7</td>
</tr>
<tr>
<td>Burstein (1980)</td>
<td>5</td>
</tr>
<tr>
<td>Raudenbush, Bryk, Cheong, Congdon, &amp; du Toit (2004)</td>
<td>5</td>
</tr>
<tr>
<td>Hox (2002)</td>
<td>4</td>
</tr>
<tr>
<td>Porter &amp; Umbach (2001)</td>
<td>4</td>
</tr>
<tr>
<td>Porter (2005)</td>
<td>4</td>
</tr>
<tr>
<td>Singer (1998)</td>
<td>4</td>
</tr>
<tr>
<td>Snijders &amp; Bosker (1999)</td>
<td>4</td>
</tr>
<tr>
<td>King (1999)</td>
<td>3</td>
</tr>
<tr>
<td>Pascarella &amp; Terenzini (1991)</td>
<td>3</td>
</tr>
<tr>
<td>Raudenbush, Bryk, Cheong, &amp; Congdon (2001)</td>
<td>3</td>
</tr>
<tr>
<td>Robinson (1950)</td>
<td>3</td>
</tr>
<tr>
<td>Berger &amp; Milem (2000)</td>
<td>2</td>
</tr>
<tr>
<td>Cohen (1988)</td>
<td>2</td>
</tr>
<tr>
<td>Cranton &amp; Smith (1990)</td>
<td>2</td>
</tr>
<tr>
<td>Hu &amp; Kuh (2003)</td>
<td>2</td>
</tr>
<tr>
<td>Kalton (1983)</td>
<td>2</td>
</tr>
<tr>
<td>Loeb (2003)</td>
<td>2</td>
</tr>
<tr>
<td>Pascarella (1985)</td>
<td>2</td>
</tr>
<tr>
<td>Patrick (2001)</td>
<td>2</td>
</tr>
<tr>
<td>Pike (2007)</td>
<td>2</td>
</tr>
<tr>
<td>Schwartz (2007)</td>
<td>2</td>
</tr>
<tr>
<td>Thomas (2006)</td>
<td>2</td>
</tr>
<tr>
<td>Weidman (1989)</td>
<td>2</td>
</tr>
<tr>
<td>Abrami, d’Appollonia, &amp; Rosenfeld (2007)</td>
<td>1</td>
</tr>
<tr>
<td>Arnold (1992)</td>
<td>1</td>
</tr>
<tr>
<td>Astin (1977)</td>
<td>1</td>
</tr>
<tr>
<td>Astin (1984)</td>
<td>1</td>
</tr>
<tr>
<td>Astin (1993)</td>
<td>1</td>
</tr>
<tr>
<td>Austin (1996)</td>
<td>1</td>
</tr>
<tr>
<td>Berger (2000)</td>
<td>1</td>
</tr>
<tr>
<td>Reference</td>
<td>Count</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td>Bock (1989)</td>
<td>1</td>
</tr>
<tr>
<td>Brewer &amp; Ehrenberg (1996)</td>
<td>1</td>
</tr>
<tr>
<td>Brewer, Eide, &amp; Ehrenberg (1999)</td>
<td>1</td>
</tr>
<tr>
<td>Bryk &amp; Raudenbush (1996)</td>
<td>1</td>
</tr>
<tr>
<td>Bryk, Raudenbush, &amp; Congdon (1994)</td>
<td>1</td>
</tr>
<tr>
<td>Bryk, Raudenbush, &amp; Condon (1996)</td>
<td>1</td>
</tr>
<tr>
<td>Burstein &amp; Miller (1981)</td>
<td>1</td>
</tr>
<tr>
<td>Busing (1993)</td>
<td>1</td>
</tr>
<tr>
<td>Chatman (2007)</td>
<td>1</td>
</tr>
<tr>
<td>Chickering (1974)</td>
<td>1</td>
</tr>
<tr>
<td>Cohen, Cohen, West, &amp; Aiken (2003)</td>
<td>1</td>
</tr>
<tr>
<td>Dale &amp; Krueger (2002)</td>
<td>1</td>
</tr>
<tr>
<td>de Leeuw &amp; Kreft (1995)</td>
<td>1</td>
</tr>
<tr>
<td>DiPrete &amp; Forristal (1994)</td>
<td>1</td>
</tr>
<tr>
<td>Draper (1995)</td>
<td>1</td>
</tr>
<tr>
<td>Duncan, Jones, &amp; Moon (1998)</td>
<td>1</td>
</tr>
<tr>
<td>Enders &amp; Tofighi (2007)</td>
<td>1</td>
</tr>
<tr>
<td>Ethington (2000)</td>
<td>1</td>
</tr>
<tr>
<td>Goldstein (1995)</td>
<td>1</td>
</tr>
<tr>
<td>Goldstein, Rabash, Plewis, Draper, Brown, Yang, …Healy (1998)</td>
<td>1</td>
</tr>
<tr>
<td>Greene (1997)</td>
<td>1</td>
</tr>
<tr>
<td>Groves, Fowler, Cooper, Lepkowski, Singer, &amp; Tourangeau (2004)</td>
<td>1</td>
</tr>
<tr>
<td>Hahs-Vaughn (2006)</td>
<td>1</td>
</tr>
<tr>
<td>Haney (1980)</td>
<td>1</td>
</tr>
<tr>
<td>Heck &amp; Thomas (2009)</td>
<td>1</td>
</tr>
<tr>
<td>Hoffman &amp; Gavin (1998)</td>
<td>1</td>
</tr>
<tr>
<td>Hox (1995)</td>
<td>1</td>
</tr>
<tr>
<td>Hu &amp; Kuh (2004)</td>
<td>1</td>
</tr>
<tr>
<td>Jonge, Gerard, Landeweerd, &amp; Nijhuis (1999)</td>
<td>1</td>
</tr>
<tr>
<td>Kennedy, Teddlie, &amp; Stringfield (1993)</td>
<td>1</td>
</tr>
<tr>
<td>Kish (1992)</td>
<td>1</td>
</tr>
<tr>
<td>Korn &amp; Graubard (1995)</td>
<td>1</td>
</tr>
<tr>
<td>Lee &amp; Bryk (1989)</td>
<td>1</td>
</tr>
<tr>
<td>Lee &amp; Smith (1997)</td>
<td>1</td>
</tr>
<tr>
<td>Lietz (1996)</td>
<td>1</td>
</tr>
<tr>
<td>Littell, Milliken, &amp; Stroup (1996)</td>
<td>1</td>
</tr>
<tr>
<td>Marsh &amp; Overall (1981)</td>
<td>1</td>
</tr>
<tr>
<td>Marsh &amp; Roche (1997)</td>
<td>1</td>
</tr>
<tr>
<td>Marsh (1980)</td>
<td>1</td>
</tr>
</tbody>
</table>
Marsh (1987) 1
Marsh (2007) 1
Marsh & Dunkin (1997) 1
Mayer (1998) 1
McLachlan & Krishnan (1997) 1
Merlo, Östergren, Lindstrom, Lindgren, & Melander (2001) 1
Morris (1995) 1
Múthen & Múthen (2004) 1
Onwuegbuzie & Daniel (2003) 1
Painter (n.d.) 1
Perna (2003) 1
Peugh & Enders (2005) 1
Pike, Kuh, & Gonyea (2003) 1
Prosser, Rabash, & Goldstein (1996) 1
Raudenbush & Bryk (1986) 1
Raudenbush & Willms (1991) 1
Raudenbush, Bryk, & Cheong (2000) 1
Raudenbush, Bryk, & Congdon (2006) 1
Rowe & Hill (1998) 1
Rowe (1999) 1
Rowe, Hill, & Holmes-Smith (1995) 1
Rugutt (2001) 1
Rust (1985) 1
Shin, Espin, Deno, & McConnell (2004) 1
Singer & Willett (2003) 1
Snijders & Bosker (2002) 1
Stapleton & Lissitz (1999) 1
Tate (2004) 1
Ting (2000) 1
Toutkoushian & Smart (2001) 1
Von Secker & Lissitz (1997) 1
Wendorf & Alexander (2005) 1
Zeegers (2001) 1
APPENDIX I

HLM STUDIES SAMPLE DATA SOURCES
<table>
<thead>
<tr>
<th>Source</th>
<th>$n^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Survey of Student Engagement</td>
<td>11</td>
</tr>
<tr>
<td>Original Data</td>
<td>10</td>
</tr>
<tr>
<td>NCES – NSOPF</td>
<td>6</td>
</tr>
<tr>
<td>HERI – CIRP First Year Survey</td>
<td>5</td>
</tr>
<tr>
<td>NCES – B&amp;B</td>
<td>4</td>
</tr>
<tr>
<td>HERI – College Students’ Beliefs and Values Survey</td>
<td>4</td>
</tr>
<tr>
<td>HERI – College Senior Survey</td>
<td>3</td>
</tr>
<tr>
<td>HERI – Faculty Survey</td>
<td>3</td>
</tr>
<tr>
<td>IPEDS</td>
<td>3</td>
</tr>
<tr>
<td>Survey of Earned Doctorates</td>
<td>2</td>
</tr>
<tr>
<td>Institutional Course Evals</td>
<td>2</td>
</tr>
<tr>
<td>Faculty Survey of Student Engagement</td>
<td>2</td>
</tr>
<tr>
<td>College Student Experiences Questionnaire</td>
<td>1</td>
</tr>
<tr>
<td>Community College Student Experiences Questionnaire</td>
<td>1</td>
</tr>
<tr>
<td>Community College Survey of Student Engagement</td>
<td>1</td>
</tr>
<tr>
<td>ETS Data</td>
<td>1</td>
</tr>
<tr>
<td>First Year Initiative Survey</td>
<td>1</td>
</tr>
<tr>
<td>HERI – Follow Up Survey</td>
<td>1</td>
</tr>
<tr>
<td>National Longitudinal Survey of Freshmen</td>
<td>1</td>
</tr>
<tr>
<td>NCES – BPS</td>
<td>1</td>
</tr>
<tr>
<td>National Orientation Directors Association Survey</td>
<td>1</td>
</tr>
<tr>
<td>Norway Faculty Survey</td>
<td>1</td>
</tr>
</tbody>
</table>

$^aN = 60$
REFERENCE LIST


223


VITA

Kimberly Q. Fath, born and raised in the Chicago area, attended Boston College and completed a Bachelor of Arts and Master of Arts degrees in Mathematics. She taught mathematics and computer programming at Boston College and Suffolk University in the Boston area. She relocated to Chicago and worked at Saint Xavier University, first as the Mathematics Specialist and later as the Director of Learning Assistance Services and the University Student Support Services Project, a TRIO program. After several years working in academic and retention support services at Saint Xavier University she enrolled at Loyola University Chicago, seeking a doctoral degree in the Higher Education program.

During her time at Loyola, Dr. Fath worked in technology support for the School of Education, helping to create the school’s first website; served as a teaching assistant and later an instructor for RMTD 400, Introduction to Research Methods, and served as campus liaison and project associate for the Multi-Institutional Study of Leadership. She has co-authored papers on the leadership development of STEM women and applications of interdisciplinary work on student learning and has presented at university, regional, and national conferences.

Dr. Fath is the Assessment Specialist in the Office of Institutional Research and Assessment at Elon University. She resides in Chapel Hill, North Carolina with her husband and son, both named Michael, and daughters Kira and Emma.