Abstract:
This study compared three college-related measures of former high school students from a reform-based mathematics curriculum, produced by the Systemic Initiative for Montana Mathematics and Science (SIMMS) Project, with those of students from more traditional mathematics curricula. The measures of interest were ACT and SAT scores, freshman mathematics grades, and initial choices of majors. The subjects entered one of two state universities in Montana in Fall 1998. An Instructor Survey of classroom practices measured the intensity of reform in college freshman mathematics classrooms, and Supervisor Interviews produced qualitative data as background.

Results from the study included evidence of tracking of former SIMMS and non-SIMMS subjects, born out by the higher nonmathematical ACT and SAT scores, mathematical ACT and SAT scores, and freshman mathematics grades of the non-SIMMS group. Conversely, SIMMS students chose science, mathematics, engineering, and technology majors more frequently than the non-SIMMS group but not significantly more frequently. Results also showed that, among three college mathematics Course Types of Developmental, Before Calculus, and Calculus, the Before Calculus courses were more reformed. However, there was no evidence to show that more reformed college mathematics courses favored either SIMMS or non-SIMMS group in terms of their grades. Interviews with 16 course supervisors affirmed the relatively traditional nature in most of the freshman mathematics courses at both Montana universities, which was quite different from what the NCTM Standards recommended. Additionally, freshman mean grades in the Before Calculus and Calculus groups were significantly higher than in the Developmental group.

The researcher concluded that 1) high schools need to monitor tracking of students; 2) the role of algebra in high school and college curricula needs to be redefined in terms of current technology and traditional value; 3) reconciliation between reformers and traditionalists must happen so that students can benefit from unified goals; 4) a new definition of academic intensity of the high school curriculum could provide focus and begin reconciliation between reformers and traditionalists; and 5) the gatekeeper nature of college developmental mathematics courses demands research into their effectiveness.
A COMPARISON OF FORMER SIMMS AND NON-SIMMS STUDENTS
ON THREE COLLEGE-RELATED MEASURES

by

Michael Allen Lundin

A dissertation submitted in partial fulfillment
of
Doctor of Education
in
Education

MONTANA STATE UNIVERSITY—BOZEMAN
Bozeman, Montana

April 2001
APPROVAL

of a dissertation submitted by

Michael Allen Lundin

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Dr. Kenneth Borland
(Signature) 4/11/01
Date

Approved for the Department of Education

Dr. Gloria Gregg
(Signature) 4/20/01
Date

Approved for the Department of Graduate Studies

Dr. Bruce McLeod
(Signature) 4/20/01
Date
STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to Bell & Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Signature

Date 4/20/01
This work is dedicated to my Mother and my Father without whose guidance I would have had neither the courage to begin nor the tenacity to finish. They are, and always shall be, my guides.
ACKNOWLEDGEMENTS

I wish to thank Dr. Ken Borland and Dr. Maurice Burke. Without their continued guidance throughout the process of planning and writing, this work would not have been possible. Through more than one storm, they held the tiller steadfast. This ship found port because of their foresight, hindsight, and insight.

I am also indebted to Dr. Rich Efoward, who offered solid perspective to the methodology of this study, and to the other members of my committee, Dr. Chris Stryker, Dr. Elisabeth Swanson, Dr. Chris Yakymyshyn, and Dr. Lyle Andersen. They endured the requisite alterations inherent in the dissertation process with grace and dignity. Their suggestions led to a better product.

Finally, without the critical insights of Lynette Felix or the gracious concerns and encouragement of Diana Paterson in the SIMMS IM office, this work could not have been completed. I am grateful for their talents, their encouragement, and their friendship.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>1. INTRODUCTION TO THE STUDY</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Problem, Purpose, and Question</td>
<td>5</td>
</tr>
<tr>
<td>Background</td>
<td>5</td>
</tr>
<tr>
<td>SIMMS vs. Traditional: A Comparison</td>
<td>5</td>
</tr>
<tr>
<td>The SIMMS Project</td>
<td>7</td>
</tr>
<tr>
<td>Research Problem</td>
<td>9</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>9</td>
</tr>
<tr>
<td>Basis for the Study</td>
<td>10</td>
</tr>
<tr>
<td>Questions of National Interest</td>
<td>10</td>
</tr>
<tr>
<td>The Perspective of Change Theory</td>
<td>10</td>
</tr>
<tr>
<td>Principles, Drivers, and Professional Standards</td>
<td>11</td>
</tr>
<tr>
<td>Research Questions, Objectives, and Variables</td>
<td>14</td>
</tr>
<tr>
<td>Primary Question</td>
<td>14</td>
</tr>
<tr>
<td>Subsidiary Question</td>
<td>14</td>
</tr>
<tr>
<td>Objectives</td>
<td>14</td>
</tr>
<tr>
<td>Variables: Freshman Data</td>
<td>15</td>
</tr>
<tr>
<td>Variables: Instructor Survey</td>
<td>16</td>
</tr>
<tr>
<td>Introduction to the Study</td>
<td>16</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td>16</td>
</tr>
<tr>
<td>A National Need to Know: The Math Wars</td>
<td>16</td>
</tr>
<tr>
<td>A Local Need to Know: High Interest: Little Data</td>
<td>17</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>18</td>
</tr>
<tr>
<td>Assumptions and Limitations of the Study</td>
<td>20</td>
</tr>
<tr>
<td>Format for the Study</td>
<td>22</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. LITERATURE REVIEW</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter Introduction</td>
<td>25</td>
</tr>
<tr>
<td>Criteria for Literature Selection</td>
<td>25</td>
</tr>
<tr>
<td>Theme Development</td>
<td>31</td>
</tr>
<tr>
<td>Context of the Problem</td>
<td>31</td>
</tr>
<tr>
<td>Problem Statement in Context</td>
<td>31</td>
</tr>
</tbody>
</table>
### TABLE OF CONTENTS - CONTINUED

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change</td>
<td>32</td>
</tr>
<tr>
<td>The Disparity Between Assessment Types</td>
<td>34</td>
</tr>
<tr>
<td>The Disparities in Program Assessment and Evaluation</td>
<td>36</td>
</tr>
<tr>
<td>The Need for Deep Change</td>
<td>42</td>
</tr>
<tr>
<td>The Evolution of Standards and NSF's Drivers of Reform</td>
<td>43</td>
</tr>
<tr>
<td>A Consequence of Mutual Adaptation</td>
<td>47</td>
</tr>
<tr>
<td>Two Levers for Reform: Curriculum and Program Evaluation</td>
<td>50</td>
</tr>
<tr>
<td>The STEP Project: A Partner in Reform</td>
<td>53</td>
</tr>
<tr>
<td>Current Understanding of the Problem</td>
<td>54</td>
</tr>
<tr>
<td>Modern Standards Mathematics Education</td>
<td>54</td>
</tr>
<tr>
<td>The Core-Plus Project: Research Meets Practice</td>
<td>56</td>
</tr>
<tr>
<td>The SIMMS Curriculum and College Freshmen</td>
<td>58</td>
</tr>
<tr>
<td>The Current State of Standards-based Practices</td>
<td>60</td>
</tr>
<tr>
<td>College-related Measures, Performance, and Graduation</td>
<td>61</td>
</tr>
<tr>
<td>Review of Previous Research</td>
<td>63</td>
</tr>
<tr>
<td>NCTM Standards-based Instruction and College Freshmen</td>
<td>63</td>
</tr>
<tr>
<td>Two Contradictory Core-Plus Studies</td>
<td>64</td>
</tr>
<tr>
<td>The SIMMS Evaluation of Student Outcomes</td>
<td>70</td>
</tr>
<tr>
<td>The Meaningfulness of College-related Measures</td>
<td>73</td>
</tr>
<tr>
<td>How Reformed Are Mathematics Professors?</td>
<td>80</td>
</tr>
<tr>
<td>Summary of Methodologies</td>
<td>87</td>
</tr>
<tr>
<td>Introduction</td>
<td>87</td>
</tr>
<tr>
<td>College-related Measures</td>
<td>88</td>
</tr>
<tr>
<td>The Core-Plus Study</td>
<td>89</td>
</tr>
<tr>
<td>The SIMMS Study</td>
<td>90</td>
</tr>
<tr>
<td>The Faculty Studies</td>
<td>92</td>
</tr>
<tr>
<td>Conclusions</td>
<td>94</td>
</tr>
<tr>
<td>Summary of the Review of the Literature</td>
<td>94</td>
</tr>
<tr>
<td>Weaknesses and Strengths in the Literature</td>
<td>96</td>
</tr>
<tr>
<td>Gaps and Saturation Points in the Literature</td>
<td>97</td>
</tr>
<tr>
<td>Avenues of Further Inquiry</td>
<td>99</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>100</td>
</tr>
</tbody>
</table>

### 3. RESEARCH METHODOLOGY

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter Introduction</td>
<td>103</td>
</tr>
<tr>
<td>Participants</td>
<td>104</td>
</tr>
<tr>
<td>Population and Subjects</td>
<td>104</td>
</tr>
<tr>
<td>Freshman Data</td>
<td>104</td>
</tr>
<tr>
<td>Instructor Survey</td>
<td>105</td>
</tr>
<tr>
<td>Supervisor Interviews</td>
<td>106</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS - CONTINUED

Supervisor Interviews..........................................................................................167
   Introduction...........................................................................................................167
   Analysis of the Responses from Supervisor Interviews.................................168
Meaning and Discussion of the Analysis of Data..................................................177
   Significance of the Results..................................................................................177
Supplemental Analysis............................................................................................181
   Relationship of the Freshman Data to the Research..........................................181
   Relationship of the Instructor Survey to the Research.......................................182
   Relationship of the Supervisor Interviews to the Research..............................183
Weaknesses, Uncontrolled Factors, and Incongruities.........................................184
Chapter Summary..................................................................................................186

5. CONCLUSION TO THE STUDY........................................................................189

   Chapter Introduction...........................................................................................189
   The Research Questions Revisited.................................................................191
   Answering the Research Questions...............................................................195
Broader Implications..............................................................................................195
   Introduction...........................................................................................................195
   Pervasive Performance Differences and Early Tracking...............................195
   Cultural Differences Between Reformers and Traditionalists.......................197
   Reformers, Traditionalists, and Mathematical Content...................................200
   What These Implications Mean..........................................................................202
Recommendations: Research and Reconciliation....................................................203
Chapter Summary..................................................................................................207

REFERENCES........................................................................................................209

APPENDICES.........................................................................................................218

   APPENDIX A.......................................................................................................219
      The Classroom Practices Inventory...............................................................220
      Freshman Course Supervisor Interview Protocol.........................................221
   APPENDIX B.......................................................................................................222
      The NSF Drivers of Reform...........................................................................222
      Principles and Standards for School Mathematics.........................................224
   APPENDIX C.......................................................................................................225
      SMET and Non-SMET Majors at University E and University W...................226
LIST OF TABLES

Table | Page
--- | ---
1. Fall 2000 Freshman Mathematics Courses Offered by Institution and Category | 109
2. Dependent Variables and Corresponding Standards for the College Instructors Survey | 110
3. Requested Fields for Freshman Data | 113
4. Timeframe for the Study | 129
5. University E: College Entrance Exam Descriptive Statistics | 134
6. University W: College Entrance Exam Descriptive Statistics | 135
7. University E: t-values and p-values For College Entrance Examinations | 136
8. University W: t-values and p-values For College Entrance Examinations | 137
9. Descriptive Statistics for Fall 1998 Grades at University E | 142
10. ANOVA Table for Fall 1998 Grades at University E | 143
11. Descriptive Statistics for Spring 1999 Grades at University E | 146
12. ANOVA Table for Spring 1999 Grades at University E | 146
13. Descriptive Statistics for Fall 1998 Grades at University W | 150
14. ANOVA Table for Fall 1998 Grades at University W | 150
15. Descriptive Statistics for Spring 1999 Grades at University W | 154
16. ANOVA Table for Spring 1999 Grades at University W | 154
17. University E: Curriculum Type vs. Major Choice for 1998 Freshmen | 158
18. University W: Curriculum Type vs. Major Choice of 1998 Freshmen | 158
19. Descriptive Statistics for Mean Responses on Instructor Surveys | 162
xi

LIST OF TABLES - CONTINUED

20. ANOVA Statistics Comparing Rscore by Institution and Course Type.........162

21. Comparative Results Between Former SIMMS and Non-SIMMS Students On College-related Performance Measures..............................197
LIST OF FIGURES

1. Mean Grade Points for Freshman Mathematics Courses at University E in Fall 1998 by Course Type and Curriculum Type..................143
2. Mean Grade Points for Freshman Mathematics Courses at University E in Spring 1999 by Course Type and Curriculum Type.................147
3. Mean Grade Points for Freshman Mathematics Courses at University W in Fall 1998 by Course Type and Curriculum Type.....................151
4. Mean Grade Points for Freshman Mathematics Courses at University W in Spring 1999 by Course Type and Curriculum Type...........................................155
5. Mean Rscores for each Course Type by Institution..............................................163
6. University E: Instructor Response Means by Course Type..................................165
7. University W: Instructor Response Means by Course Type.................................165
8. SIMMS Ideal vs. Combined Instructor Practices by Course Type......................199
ABSTRACT

This study compared three college-related measures of former high school students from a reform-based mathematics curriculum, produced by the Systemic Initiative for Montana Mathematics and Science (SIMMS) Project, with those of students from more traditional mathematics curricula. The measures of interest were ACT and SAT scores, freshman mathematics grades, and initial choices of majors. The subjects entered one of two state universities in Montana in Fall 1998. An Instructor Survey of classroom practices measured the intensity of reform in college freshman mathematics classrooms, and Supervisor Interviews produced qualitative data as background.

Results from the study included evidence of tracking of former SIMMS and non-SIMMS subjects, born out by the higher nonmathematical ACT and SAT scores, mathematical ACT and SAT scores, and freshman mathematics grades of the non-SIMMS group. Conversely, SIMMS students chose science, mathematics, engineering, and technology majors more frequently than the non-SIMMS group but not significantly more frequently. Results also showed that, among three college mathematics Course Types of Developmental, Before Calculus, and Calculus, the Before Calculus courses were more reformed. However, there was no evidence to show that more reformed college mathematics courses favored either SIMMS or non-SIMMS group in terms of their grades. Interviews with 16 course supervisors affirmed the relatively traditional nature in most of the freshman mathematics courses at both Montana universities, which was quite different from what the NCTM Standards recommended. Additionally, freshman mean grades in the Before Calculus and Calculus groups were significantly higher than in the Developmental group.

The researcher concluded that 1) high schools need to monitor tracking of students; 2) the role of algebra in high school and college curricula needs to be redefined in terms of current technology and traditional value; 3) reconciliation between reformers and traditionalists must happen so that students can benefit from unified goals; 4) a new definition of academic intensity of the high school curriculum could provide focus and begin reconciliation between reformers and traditionalists; and 5) the gatekeeper nature of college developmental mathematics courses demands research into their effectiveness.
INTRODUCTION TO THE STUDY

Chapter Introduction

Examined in this study was the question of whether two groups of Montana university freshmen differed with respect to college-related performance. Operationalizing this question meant comparing ACT and SAT test scores, freshman mathematics grades, and initial choices of majors for both groups. Comprising one group of subjects were students from a reform-based high school mathematics curriculum, produced by the Systemic Initiative for Montana Mathematics and Science (SIMMS) Project. In the alternate group were students from more traditional high school mathematics backgrounds. The students in this study entered one of the two state universities in Montana in 1998, and each site is treated separately. This study also included a survey of pedagogical practices in freshman mathematics classrooms at the same two Montana universities, as well as interviews with freshman mathematics course supervisors at those institutions. The survey and the interviews provided data for an informational framework, based on reform practices in freshman mathematics classes, instrumental in explaining potential differences in grades between the SIMMS and non-SIMMS students.

It was a premise of this study that comparing the performances of students from traditional and reform-based curricula, such as SIMMS, was not a trivial matter. Several facts, supported in the review of the literature, buttressed this premise. First, traditional
goals and prospective outcomes for learners differed from those of reform-based curricula, as did their respective theoretical foundations (Romberg & Wilson, 1995, p. 4). Therefore, assessment measures related to either type of curriculum alone, while useful for gauging performance for that particular curriculum, were not valid when comparing the two types.

Mathematics departments at both state universities in Montana, for example, used ACT and SAT scores or placement tests to sort students into their first university mathematics courses. However, a traditional test, such as the ACT, SAT, or a placement exam, does not and cannot measure ability outside its validity, and neither traditional mathematics content nor traditional test-taking processes align well with the content or processes valued by mathematics reformers (Romberg & Wilson, 1995, p. 4).

Second, university mathematics instructors practiced traditional teaching methods (LaBerge, Zollman, & Sons, 1997, p. 13) different from what many students may have experienced in high school. For SIMMS students in particular, more learner-focused interaction in high school (Dapples, 1995, p. 77; Thompson, 1992, p. 136) may have been replaced by lectures in college. Moreover, university faculty usually control their curricula (Janzow, Hinni, & Johnson, 1997, p. 499), including course assessments and evaluations. Consequently, former SIMMS students, caught between traditional and reform paradigms, found disturbing the traditional nature of the college mathematics classroom (Allinger, Lott, & Lundin, 1998, p. 29).

Third, reform-based standards, the products of mathematics professional organizations at secondary and post-secondary levels, were evolving during the period when subjects were in high school. While these various sets of guidelines seemed to
agree in their goals and practices for mathematics education, they were relatively novel. They were still changing, and all mathematics educators did not embrace them (LaBerge, Zollman, & Sons, 1997, p. 6).

Assessing the performance of SIMMS and non-SIMMS students in an equitable manner, then, was a matter of compromise. As such, ACT and SAT scores and freshman mathematics grades were of interest, since SAT and ACT scores correlate well with college freshman grade point averages (College Entrance Examination Board, 2000; ACT, Inc., 2000). Also, freshman course grades are "the single most revealing indicator of [students'] successful adjustment to the intellectual demands of a particular college's course of study" (Pascarella & Terenzini, 1991, p. 388). Importantly, data from both measures were readily available, since a concordance between ACT and SAT tests is often used to make decisions about admissions and placement (Dorans, Felicia, Pommerich, & Houston, 1997, p. 24). The traditional nature of ACT and SAT tests was of concern, however, because constructs measured by them may have underplayed the abilities of SIMMS students, whose experience had not been traditional. Hence, to better explain potential performance differences between former SIMMS and non-SIMMS students, an Instructor Survey produced data on practices in freshman mathematics classes. An analysis of this data provided some information about reform practices useful in explaining potential differences in freshman grades.

In addition to ACT and SAT scores and freshman grades, a third measure in this study, an initial choice of majors, was important. This measure represented student interest in a particular academic field, influenced, primarily, by pre-college experiences. This measure was chosen, then, to uncover differences in aspirations between former
SIMMS and non-SIMMS students. Especially, students' choices of majors in science, mathematics, engineering, and technology (SMET) were of interest here.

Since the reform movement is an embodiment of change, the theory of change entered into the subsequent exposition as a natural framework, without which, results would have been less meaningful. Additionally, various sets of professional standards stipulated the values, beliefs, and practices that might have guided mathematics instruction in a reform sense, and as such, are espoused by secondary and post-secondary professional organizations. That those standards have evolved and are still evolving with the reform movement is an idea that merited examination, if any recommendations were to stand firm as the bedrock of best practice. That evolution was examined here.

In conclusion, ACT and SAT scores, freshman grades, and initial choices of majors were measures that were important to Montana students (and others) in their university experiences. These measures represented available data that were used to compare the performance of former SIMMS and non-SIMMS students. In an academic arena challenged by the process of reform, the instructor survey of classroom practices and the interviews with freshman mathematics course supervisors provided background about the reform practices in Montana university mathematics classrooms. Differences in performance became more meaningful within the context of change theory, driven by various reform-based tenets of professional mathematics and collegiate organizations.

This chapter has three remaining parts. In the first part the research problem, purpose of the study, basis for the study, and research question are specifically addressed, beginning with brief history of the problem. In the second part of this chapter is a comprehensive introduction to the study, including its significance, operational
definitions, and assumptions and limitations. A chapter summary follows these two main sections.

**Problem, Purpose, and Question**

**Background.**

**SIMMS vs. Traditional: A Comparison.** The SIMMS curriculum materials embody the National Council of Teachers of Mathematics (NCTM) Standards (NCTM, 1989) for grades nine through twelve in a "learner-focused" (Thompson, 1992, p. 136) model of pedagogy. Each module begins with an Exploration or Activity that is built around a real-world problem. Embedded in the problem-solving process is the use of calculators or computers, and each lesson emphasizes cooperative learning by way of a discussion component. The use of physical materials or models is an integral part of many lessons, as are assessments that involve writing about mathematics. Students are encouraged to give presentations of work in progress and presentations of completed projects (MCTM, 1996-1998). These materials, in contrast to those that are more traditional, encourage a method of learning and teaching of mathematics that is anything but "linear subject, mainly concerned with mechanistically teaching facts and skills predominately related to number and generally characterized by paper-and-pencil activity..." (Nickson, 1992, p. 103).

On the other hand, Thompson (1992) would call that traditionalist approach "content focused" (p. 136), driven by the teacher-centered attempts to present content and explain it to students in order to foster conceptual understanding. A content-focused
approach may have an emphasis on understanding or an emphasis on performance, the first, driven by the structure of mathematics itself, and the second, driven by an instrumental or pedagogical approach. In either case, however, the mathematical content, more than the learners' abilities or interests, guides practice.

Consider, for example, the concept of linear equations, often a content item in high school freshman mathematics courses. A SIMMS approach to the subject, "Are You Just a Small Giant" (Carspecken, Eichenburger, Johnson, & Souhrada, 1997) begins with a brief reading about the world's tallest human, which is followed by formal definition of the geometric concept of similarity. This, in turn, is followed by a discussion of proportionality of various body parts. Students are guided through a lesson in which they measure each classmate's foot and compare this to his or her height. Gradually, they develop a natural constant that is the slope of a linear model of height as a function of foot size. Written activities and discussions that stem from the concepts of proportionality, growth patterns, and related subjects follow this lesson. The lesson, thereby, integrates activities, cooperative learning, technology, mathematical notions, and scientific topics in a manner consistent with the NCTM Standards. Notably, The Classroom Practices Inventory (APPENDIX A), used in this study to survey freshman mathematics instructors, captured the essence of those practices embedded in the SIMMS materials. Of course, the separate questions on that instrument were designed to measure practices recommended National Council of Teachers of Mathematics.

In contrast, a typical content-focused approach formally depends on definitions of proportions, slopes, linear functions, and similarity. Students might be given data but would not participate in its collection or the development of any model. Instead, a
mathematical model might be provided for them, and much of their subsequent work is analytic rather than synthetic. Any applications to science are secondary to the formal mathematical notion of linear function, and technology might not be a part of the process.

The literature review uncovered similarities between the NCTM Standards and other documents that purported to guide mathematics instruction. The Standards for Introductory College Mathematics Before Calculus (Cohen, 1995) and in the Guidelines for Programs and Departments in Undergraduate Mathematics (MAA, 1998a), for example, prescribed how college and university instructors should teach. More generally, The Principles of Good Practice for Undergraduate Education (Chickering & Gamson, 1987) encouraged active learning, cooperative learning, and diverse ways of learning. Therefore, much of what was prescribed in all of these documents was found in the SIMMS modules, and those practices were reflected in the Classroom Practice Inventory in this study.

In summary, substantial differences existed between traditional and SIMMS curricula. Those differences naturally led to questions about the comparative performance of former SIMMS and non-SIMMS students on college-related measures, and those questions have guided this study.

The SIMMS Project. In 1991 the National Science Foundation (NSF) funded the Montana Council of Teachers of Mathematics with a ten million-dollar grant, enabling the Systemic Initiative for Montana Mathematics and Science (SIMMS) Project to begin five years of operation (The SIMMS Project, 1996, p. 1). A major goal of the SIMMS Project was to write, publish, and put into practice a mathematics curriculum for grades
nine through twelve that would reflect the (then) recently published *Curriculum and Evaluation Standards for School Mathematics* (NCTM), 1989).

As planned in the grant proposal, the SIMMS Project implemented its curriculum, based on the NCTM Standards, in many Montana high schools (MCTM, 1997, p. 6). Four years later, in the fall of 1996, the first cohort of SIMMS-educated mathematics students began matriculating at Montana's institutions of higher education. Although the SIMMS Project had evaluated its pilot curriculum at the high school level (MCTM, 1998), it had not yet studied the effects of its reform-based curriculum on college students, since that cadre of subjects had not graduated from Montana high schools. Realizing that "Preparing students for post-secondary education involves more than simply guiding [students] through SAT or ACT tests and the admissions process" (Allinger, Lott, & Lundin, 1998, p. 16), the SIMMS Project staff conducted pilot research to better understand the attitudes and performance of college freshman who had been SIMMS students. That research was timely, having been prompted by public and professional interest in the college performance of SIMMS veterans, as communicated by Glenn Allinger, SIMMS Professional Development Co-Chairman (personal communication, March 28, 2000).

How did SIMMS educated students perform in college mathematics when compared with non-SIMMS students? Analyses of qualitative and quantitative data gathered in the 1997 pilot study began to address that question (Allinger, Lott, & Lundin, 1998), and the findings were noted in the literature review. Those findings, though meager, suggested possible differences between SIMMS and non-SIMMS groups.
However, the general question of the relationship between the SIMMS curriculum and college students' performance and initial choices of majors remained open.

Research Problem

It was not known whether former SIMMS and non-SIMMS students differed in terms of college-related measures. In particular, it was not known whether they differed with respect to ACT scores, SAT scores, college freshman mathematics grades, or initial choices of major.

Purpose of the Study

The purpose of this study was to determine if variations in college-related measures existed between two groups of former Montana high school students, those from a more traditional mathematics programs and those from the SIMMS integrated mathematics program. ACT scores, SAT scores, and freshman mathematics grades were the more traditional measures used to compare the SIMMS and non-SIMMS groups, and initial choice of majors, while not a performance indicator, was a measure of interest in mathematics or a related field. The literature review lent support to this choice of measures as meaningful to the study at hand. That review also supported a survey of classroom practices of college mathematics instructors and interviews with freshman mathematics course supervisors to provide a backdrop for the discussion of performance comparisons between former SIMMS and non-SIMMS students.
Basis for the study

Questions of National Interest. Although this study was specific to Montana, it was one that sprang from much national interest in mathematics reform, about a decade old, if dated from the first publication of the NCTM Standards. However, relatively little data existed on the effects of reform-based high school mathematics on college students. The dearth of literature existed despite much funding by NSF in an effort to promote reform in mathematics and science (NSF, 1999). Furthermore, the evolution of the reform movement had spawned many general questions, begging for extensive research related to the parameters of this study. For example, the following questions were of public concern and related to the questions that guided this study.

How do students educated with reform-based curricula in general compare with those who have studied in more traditional settings? How do high school mathematics departments best prepare their students for college mathematics courses? What are effective ways to formulate curricular goals and align them with practice in both secondary and post-secondary mathematics departments? What are effective strategies to promote dialog between secondary and post-secondary mathematics departments? To what extent do mathematics departments in high schools and colleges now conform to professional standards for educating students? (Green, 1999; Mervis, 1998)

The Perspective of Change Theory. Change theory has much to do with many aspects of reform, which serves as a rich example of the dynamics in the process of change. For this study, it was necessary to enter the realm of program comparison. Viewing mathematics programs from a perspective of change theory brought forth issues,
structures, and processes associated with these programs, allowing them to be better scrutinized and compared with one another. From these comparisons came suggestions for more effective programs as well as answers to questions about differences between SIMMS and more traditional curricula.

Without this backdrop of change theory, any findings of this study would have been less meaningful, because reform goals were evolving, and those goals tended to drive mathematics assessment. Due to the dynamic nature of mathematics reform, the reform goals represented, metaphorically, moving targets. That is, for many levels of assessment, the goals changed, exemplifying what Hall and Hord (1987) called "mutual adaptation" (p. 116) of beliefs and practices of various players. For this study, mutual adaptation explained, for example, why few assessments of reform projects included formal evaluations at the post-secondary levels. Also exemplifying mutual adaptation, an initial choice of majors by college students was once recommended indicator of systemic reform by NSF (MCTM, 1997, p. 2), although it was not specifically listed later (National Science Foundation, 2000). Mutual adaptation, then, is itself evidence that mathematics reform goals are evolving.

**Principles, Drivers, and Professional Standards.** In this study, four main sets of documents had prescriptions for mathematics educators, fortified by general principles for college and high school teaching. Comprising those sets of documents were the three-volume NCTM Standards (NCTM, 1989; NCTM, 1991; NCTM, 1993), the American Mathematical Association of Two-Year Colleges (AMATYC) Crossroads in Mathematics: Standards for Introductory College Mathematics Before Calculus (Cohen,
1995), the Mathematical Association of America's (MAA) Guidelines for Programs and Departments in Undergraduate Mathematical Sciences (MAA, 1998), and the Principles for Good Practice in Undergraduate Education (Chickering & Gamson, 1987).


At the post-secondary level, the AMATYC and MAA documents served as guidelines for mathematics educators. It should be noted that the MAA Guidelines were less prescriptive than were the AMATYC Standards, which were similar to the NCTM Standards in structure and wording. More generally, similar prescriptions for mathematics educators existed among the NCTM, AMATYC, and MAA documents, and those were emphasized in the review of the literature.

Noteworthy also were the best-practice principles developed for general college teaching by Chickering and Gamson (1987), because they shared common ground with those practices recommended by the NCTM, AMATYC, and MAA. According to Chickering and Gamson in the "Seven Principles for Good Practice in Undergraduate Education," good practice encourages contact between students and faculty, collaboration among students, active learning, prompt feedback, time on task, high expectations for students, and diverse but equitable ways of knowing (Chickering and Gamson, 1987, p.
7). The deployment of cooperative learning, active forms of learning, and equitable ways of learning were three of the recommendations also found in the NCTM, AMATYC, and MAA documents. (In the MAA Guidelines see section "3. Mathematics and General Education" on p. 8.) The Instructor Survey in this study was in part concerned with two of these principles, cooperative and active learning, and with standards from the other documents as well.

Additionally and importantly, at the state and regional levels, NSF's "Drivers of Reform" were designed to guide grant awardees in the direction of systemic change. (See APPENDIX B for the NSF Drivers of Reform.) Particularly, driver five was important to this study, since in its original form, it recommended counting college majors as an indicator of systemic change (Montana Council of Teachers of Mathematics, 1997, p. 1; National Science Foundation, 2000). More about the drivers and their apparent changes was stated in the review of the literature.

In conclusion, the various sets of educational guidelines shared goals for high school and college classrooms and were important in grounding this study. Those guidelines came from professional organizations and served to steer mathematics teaching and learning down the road of reform. Together with the NSF Drivers, those sets of guidelines provided a rationale for examining an initial choice of majors among university freshmen, and for measuring the reform-based practices in freshman mathematics classes. In general they buttressed the theory and practice of mathematics reform in secondary and post-secondary classrooms, and in particular they supported that theory and practice associated with the SIMMS curriculum. For the research at hand,
those documents provided reasons to monitor initial choices of majors among college freshmen and to profile the pedagogical practices in their mathematics classes.

Research Questions, Objectives, and Variables

Primary Question. How did former SIMMS and non-SIMMS students at either Montana state university in 1998 compare with respect to three college-related measures: ACT and SAT scores, freshman mathematics course grades, and initial choices of major?

Subsidiary Question. Did the reform-type classroom practices of freshman mathematics instructors differ by Course Type: Developmental, mathematics Before Calculus, or Calculus, and if so, were these differences evident in the performance of former SIMMS and non-SIMMS students, according to their freshman mathematics grades disaggregated into the same three categories?

Objectives. There were three research objectives. The first objective involved gathering data from those college freshmen entering the two Montana state universities in Fall 1998. This study compared former SIMMS students and non-SIMMS students with respect to ACT and SAT scores, performance in freshman mathematics courses, and initial declarations of majors. An analysis demanded a sufficiently large sample of former SIMMS students, which only became possible with the entering freshman class of 1998. The comparison of SIMMS and non-SIMMS students focussed on the traditional-aged freshmen cohorts at both universities, and the data included mathematics grades from both semesters of the students' freshman year. In what follows, data from objective one was referred to as "Freshman Data."
Objective two involved surveying college mathematics instructors at the freshman level as a basis for comparing the mathematics performance of former SIMMS and non-SIMMS students. Objective two was referred to as the "Instructor Survey," and data came from the instructors' self-reports about the intensity of their classroom practices as they aligned with those recommendations of the NCTM, MAA, and AMATYC.

Objective three involved interviewing freshman mathematics course supervisors to ascertain course goals, expectations, standards, and outcomes. Data from these interviews added depth and dimension to the instructor survey data. Objective three was referred to as the "Supervisor Interviews."

Together, the results from Freshman Data, the Instructor Survey, and Supervisor Interviews provided a rational means of accounting for differences in freshman mathematics grades between SIMMS and non-SIMMS students. In addition, the results from the Instructors Survey and Supervisor Interviews described the state of reform among freshman instructors at Montana's two state universities.

Variables: Freshman Data. The independent variable for this research piece was Curriculum Type, either non-SIMMS or SIMMS, used by the respective groups of students. There were three categories of dependent variables. Entering freshman ACT and SAT scores constituted the first category. Separately, those variables were ACT Mathematics, Science Reasoning, English, Reading, and SAT Mathematics and Verbal scores. Freshman mathematics course grades, disaggregated by course type, constituted the second category of dependent variable. Lastly, the levels of SMET, non-SMET, or Undeclared, determined the third category of dependent variable, students' choices of
majors. That variable represented one of three curricular choices: science, mathematics, engineering, or technology (SMET), a major that is none of the previous (non-SMET), or the non-declaration of a major.

Variables: Instructor Survey. The independent variable for objective two of this research study, the Instructor Survey, consisted of the Course Type, having three levels: Developmental, mathematics Before Calculus, or Calculus. The dependent variables were measures consistent with professional standards for high school and college mathematics: problem-solving, cooperative learning, connections with other disciplines and other areas of mathematics, use of technology, equity, and mathematical reasoning. There were twelve variables in all, and three of these represented measures that are more traditional: working problems from a text, lecturing in class, and using worksheet-based exercises in class. See the Classroom Practices Inventory in APPENDIX A.

Introduction to the Study

Significance of the Study

A National Need to Know: The Math Wars. Politically, the mathematics education community has been involved in what has been dubbed "math wars" (Green, 1999, p. 1), an ideological battle between traditionalists and reformers. Skirmishes occurred in secondary and post-secondary settings, in school board meetings, and at state and national levels. Respected mathematics journals, such as The American Mathematical Monthly, devoted editorial pages to both camps in an attempt toward reconciliation, as indicated in recent articles by Tucker (1999) and Krantz (1999). Green
(1999) and the January edition of Focus (2000) noted that, among those taking sides in the war, were several Nobel laureates and the Secretary of Education, Richard W. Riley, joined by of educators, parents, students, and community members.

Spurred on by the findings of a Department of Education Expert Panel, formed by Secretary Riley, traditionalists wrote a letter attacking the panel's choice of ten recommended "exemplary" and "promising" curricula, all, more or less, based on the tenets of reform found in the NCTM Standards. The letter to Secretary Riley, published November 18, 1999 as a paid advertisement in the Washington Post (NCTM, 1999), was found at the Mathematically Correct web site (http://www.mathematicallycorrect.com), while the NCTM's formal support of Secretary Riley's endorsement (Thorpe, 1999) was on-line at the NCTM web site (http://www.NCTM.org/rileystatement.htm).

In spite of the polemic fueled by ideological differences between reformers and traditionalists, any real clarity of issues must occur in the trenches of research, as in this study, which compared traditional curricula to one that is more reformed. Results from this study led to suggestions for more productive secondary and post-secondary mathematics programs and to smoother transitions for incoming college freshmen.

**A Local Need to Know: High Interest, Little Data.** Only eight years have passed since the National Science Foundation awarded the original twenty-six State Systemic Initiative grants, which included the SIMMS Project, and a little more than ten years, since the NCTM published its Curriculum Standards. The first SIMMS cohort graduated from college with the class of 2000. The mathematics reform movement weathered a hearty gale in a sea of politics (Mervis, 1998; Schontzler, 2000), yet there was little
information about the relationship between the SIMMS curriculum and college performance. More surprisingly, despite national involvement and interest in reform, there was little data on the effects of any high school mathematics reform curriculum on college students, although the results of mathematics reform at the college level are beginning to surface (Hurley, Koehn, & Gantner, 1999). This was a time for research, since interest was high, and the subjects were accessible.

The SIMMS materials comprised all or a substantial part of the curriculum in 17 high schools across Montana and supplemented the curricula in many more schools nationwide, according to SIMMS staff member, Lynette Felix (personal communication, April 11, 1999). The exact number of additional schools that used SIMMS materials as supplements was not known. Only about 100 former SIMMS students at each university in this study met the criteria of three years or more of SIMMS preparation to be included in the SIMMS group. Nevertheless, more than one seventh of Montana’s 114 public high schools (OPI, 1998, p. 1) used SIMMS materials at some level at the time of this study, according to Felix, and Montana tax dollars supported that choice. Consequently, Montana citizens deserved to know as much as possible about that curriculum. This study answered several questions about the SIMMS curriculum and its relationship to college performance within the contexts of mathematics reform, change, and professional standards.

Definition of Terms

1. ACT Score: score on the American College Testing Service exam, purportedly measuring curriculum-based educational development, higher-order thinking skills,
and knowledge in college preparatory courses (Jeff Schiel, Senior Research Associate to ACT, personal communication, January 28, 1999).


3. Concordance: statistical relationship between the SAT and ACT scores allowing conversion from one type of score to another (Jeff Schiel, Senior Research Associate to ACT, personal communication, January 28, 1999). Note that no psychometric relationship between the two tests is implied, because the tests measure different constructs.


5. GPA: grade point average, based on a four-point scale.

6. Mathematics Program (high school): a full course of study of at least three years of high school mathematics (three Carnegie units) acceptable for entrance into Montana universities.

7. Mathematics Program (university): mathematics programs intended for the first two years of university learning, including developmental courses.

8. Program Evaluation: The process of determining the effectiveness of an educational program in achieving its goals and, therefore, its value in comparison to its required resources (NCTM, 1995, P. 90).

9. SAT Score: score on the Scholastic Aptitude Test I (SAT I) written by the Educational Testing Service, purportedly measuring the general ability to reason.
quantitatively (Jeff Schiel, Senior Research Associate to ACT, personal communication, January 28, 1999).

10. SMET: science, mathematics, engineering, and technology.

11. SSI: State Systemic Initiative awards by the National Science Foundation.

Assumptions and Limitations of the Study

1. The data sets received from the two Montana state universities were both complete and valid. All data received were scanned and cross-checked for accuracy with smaller data samples used in previous studies. However, it is possible that errors in data retrieval could have compromised the validity of the data.

2. Each SIMMS and non-SIMMS subject received at least three years of high school mathematics, as required for admission to both Montana state universities and no more than four years of high school mathematics. It is a possibility that subjects in the SIMMS or the non-SIMMS groups may have received more or less mathematics instruction than was required for entrance into these institutions. In this study no documentation, experimental control, or statistical control was employed to explain variation due to this phenomenon.

3. Historical and qualitative evidence was presented in the literature review to argue that substantial differences exist between the SIMMS curriculum, as it was implemented in the years 1992-1998, and the more traditional high school mathematics curricula of that time period. However, gathering extensive data to compare and contrast the many mathematics curricula of that period was not in the scope of this study. It is possible that other curriculum materials in use at that time shared characteristics
similar to those of SIMMS and that these similarities could detract from the statistical analysis of the data and conclusions of this study. It is also possible that a comparative study of those curricula could enhance the understanding of factors affecting the dependent variables in this study.

4. In this study, for the purpose of statistical analysis of freshman mathematics grades, data were disaggregated into three categories: Developmental, mathematics Before Calculus, and Calculus. It is assumed that courses in each category shared enough characteristics to warrant inclusion into that category. No statistical analysis was presented to argue for this division, which constituted a first step in course placement rubric at both universities in this study. It is possible that a finer partition of courses could have yielded more telling results, but limited sample sizes in each of the resulting categories might also have prevented valid statistical analysis.

5. Attending college is done by self-selection rather than by random assignment (Pascarella and Terenzini, 1991, p. 566). By choosing to study those freshmen students at two universities, the researcher could not account for those who went elsewhere or those who did not go to college. It is possible that not including those who choose out-of-state schools affected the mean SAT-ACT concordance or otherwise biased the values of the dependent variables measured here. It was also possible that limiting the populations to two Montana universities biased grade point averages and SMET tallies. There was no evidence, however, that the SIMMS and non-SIMMS students who had chosen to go to colleges outside Montana were disproportionate.
Format for the Study

This study consists of four more chapters in addition to this introductory chapter. The review of the literature in Chapter Two supported the context and the current understanding of freshman college-related performance, initial choices of majors, and college mathematics classroom practices as measures. The supporting literature was summarized and examined for weaknesses and strengths, and avenues of further inquiry were noted. In Chapter Three, the methodology section included information about the populations studied, the instruments and materials used for the study, and the research design. Assumptions and limitations of the methodology were included there also, as was a timeframe for the study. In Chapter Four the results of the data analysis and a discussion of the meaning of those results were included as well as a Supplemental analysis. Finally, in Chapter Five a discussion of the broader theoretical and practical implications of this study as well as its limitations can be found. Recommendations for further research, in which the research question regarding differences between SIMMS and non-SIMMS college-performance and their curricular choices, were summarily included in Chapter Five.

Chapter Summary

Challenged by the new NCTM Standards and funded by NSF, the SIMMS Project created a high school mathematics curriculum and implemented that curriculum in Montana, as the mathematics reform movement evolved. Ideological conflicts, predicted by change theory, raised awareness of the issues surrounding reform and generated...
demands for research on reform. Disparities existed between traditionalists and reformers about what constitutes curriculum, practice, and assessment of mathematics in high school and college, partially because the relationships between reform curricula and performance had not been well-documented. Also, disparities existed between what was recommended by professional standards for secondary and post-secondary mathematics programs and what was actually practiced. Finally, the evolution of mathematics reform created assessment goals that continued to evolve, confounding valid performance comparisons between former SIMMS and non-SIMMS students.

The main objective of this study was to compare the performance and initial choice of majors of incoming university freshman from two different groups: former SIMMS and non-SIMMS students. Each of the two Montana state universities was treated separately in the study. ACT and SAT scores, university freshman mathematics grades, and initial choices of majors of former SIMMS and non-SIMMS students comprised the measures here. Additionally, the Instructor Survey of classroom practices and course Supervisor Interviews provided information about university mathematics reform with which freshman grades were compared to explain differences between groups. Also, the instructor survey provided a measure of the disparity between what was recommended by professional standards and what was actually practiced in the university mathematics classroom at Montana's state universities.

The combination of traditional performance measures, ACT and SAT scores and freshman mathematics grades, and two reform-based measures, initial choice of major and the amount of reform in mathematics classes, represented the union of traditional and reform constructs. Such a union was needed to equitably compare both SIMMS and non-
SIMMS students. The measures and their associated instruments represented a compromise between accessibility of the data, validity of measured constructs, and value to both traditionalists and reformers. The results of this study began to clarify the relationship between traditional and reform mathematics in terms of high school and college performance, a relationship that the existing literature did not clarify.
LITERATURE REVIEW

Chapter Introduction

This literature review consists of four main parts: Criteria for Literature Selection, Theme Development, Conclusions, and Chapter Summary. The Themes section includes Problem Context, Understanding of the Problem, Previous Research and Findings, and a Review of Methodologies. The reader will note that problem context is complicated, and the section that describes it is necessarily detailed. The concluding section summarizes the review, points out strengths and weaknesses of the research, points to gaps and saturation levels in the knowledge base, and alludes to avenues of further inquiry.

Criteria for Literature Selection

The literature pertinent to this study was drawn from six main sources: 1) change theory; 2) guidelines for mathematics education, developed by professional organizations; 3) sources on learning, teaching, and the college curricula; 4) documents produced by the SIMMS Project; 5) research on reform in mathematics education; and 6) studies about the meaningfulness of high school and college-related performance measures.

In the literature on change theory and practice, books by Fullan and Stiegelbauer (1991) and Hall and Hord (1987) complement each other as accepted texts on planned
change. While the first text focused on the effects of change in elementary and secondary school settings, the second more formally outlined the structure and process of planned change from the perspective of the Concerns Based Adoption. The Concerns Based Adoption Model (CBAM), "provide [ed] ways to label change process phenomena, to take positive action in facilitating change, and to predict effects" (Hall & Hord, 1987, p. viii).

At the elementary, secondary, and post-secondary educational levels, sets of professional guidelines for mathematics education were written to inform practice. The pedagogical aspirations of various professional organizations were embodied in these guidelines, and they shared common prescriptions for educators. These sets of guidelines were of interest in this study, since the SIMMS curriculum (and other curricula) had roots in them, although the extent to which college mathematics instructors follow them at Montana universities was not known. Indeed, this was a question considered in the work at hand.

The three volumes of professional standards, published by the NCTM, constituted the primary framework for the mathematics reform movement. These volumes included prescriptions for curriculum and evaluation, pedagogy, and the assessment of elementary and secondary mathematics. A new volume of revised NCTM guidelines, Principles and Standards for School Mathematics (NCTM, 2000), which combined the previous three volumes with input from various professional organizations, was also of interest. After two years in draft form (NCTM, 1998), the published volume appeared in April 2000.

Documents analogous to the NCTM Standards, meant to guide post-secondary mathematics educators, were cited also. The American Mathematical Association of

Providing a more general prescription for best practice were the *Seven Principles for Good Practice in Undergraduate Education*, developed for college teaching by Chickering and Gamson (1987). According to those researchers, good practice encourages contact between students and faculty, collaboration among students, active learning, prompt feedback, time on task, high expectations for learners, and diverse but equitable ways of knowing. (Chickering and Gamson, 1987, p. 7) Similarities in prescriptions for educators between the *Principles* and the NCTM, AMATYC, and MAA documents were clearly evident, and the principles themselves were grounded in Chickering's psychosocial research and theories (Evans, Forney, & Guido-DiBrito, 1998, p. 16; Pascarella & Terenzini, 1991, p. 19).

More specifically, cooperative learning, active forms of learning, and equitable ways of learning were three of Chickering's and Gamson's principles also advocated in the NCTM, AMATYC, and MAA documents. (In the MAA Guidelines, see section "3. Mathematics and General Education" on p. 8.) Additionally, Chickering's psychosocial theories affirmed the importance of measuring cooperative and active learning in the study at hand. Pertinent to this research study, the Instructor Survey contains questions regarding two of these principles, cooperative and active learning.
Texts by Pascarella and Terenzini, (1991), Stark and Latucca (1997), and Evans, Forney, and Guido-DiBrito (1998), documenting the effects of college on students and knowledge about college curricula, were cited in this study because these texts are accepted references on the college experience. Citations from various chapters in the anthology, *Handbook of the Undergraduate Curriculum* (Gaff and Ratcliff, 1997), appeared here for the same reason. Likewise, the work of Romberg and Wilson (1995) was taken to heart in this study, because the authors exhibited extensive knowledge about assessment and mathematics education.

Much particular information for the work at hand came from the many publications of the SIMMS project. The SIMMS curriculum, the source of the phenomenon under investigation, included six levels of materials with either three or four books of modules at each curricular level (MCTM, 1996-1998). In addition to the curricular materials, the SIMMS Project staff produced a number of monographs that document the project's goals, research and evaluation studies, and a final report. Those monographs were an indispensable source of data, while the structure of the curriculum materials themselves constituted strong evidence that former SIMMS students underwent a unique educational experience.

Also of related interest was research that profiled the beliefs and practices of high school teachers (Weiss, Upton, and Nelson, 1992; Weiss, Matti, and Smith, 1994) and university faculty (LaBerge, Zollman, and Sons, 1997) with respect to recommendations of the NCTM Standards. The Classroom Practice Inventory (APPENDIX A), the Instructor Survey instrument used in this study to measure alignment of classroom practices with the NCTM Standards, came directly from the work of the latter three
researchers, having been adapted (p. 6) from a study of Weiss, Upton, and Nelson (1992). Recent dissertations by Kull (1996) and Hawkins (1998), respectively, brought to light the degree of awareness of the NCTM Standards in preparatory programs in colleges and universities and the effect of NCTM standards-based education on the self-concept and achievement of university freshmen.

Two conflicting evaluations of the Core-Plus Mathematics Program (CPMP), a reform-based high school mathematics project and a contemporary of SIMMS, were of special interest because of similarities between Core-Plus and SIMMS. Core-Plus' own evaluation, Contemporary Mathematics in Context: Evaluation Results, included an examination of college student performance and is written and printed for public consumption. (Core-Plus Mathematics Project, 2000b). The glossy report contrasted in substance, appearance, and results with an independent study of college student performance prepared by two mathematicians, Bachelis (1998, 1999) and Milgram (1999). Those researchers worked in tandem to investigate the achievement and attitudes of college freshmen and former CPMP students. Like the Core-Plus evaluators, Bachelis and Milgram compared SAT and ACT scores, freshman mathematics grades, course placement, and attitudes of former CPMP students and students of more traditional curricula. Unlike Core-Plus' findings, however, Bachelis' and Milgram's analysis cast CPMP in unfavorable light. They self-published their findings, and because both researchers are mathematicians, their work, because of its implications and political ramifications, had an impact that could not be dismissed. Notably, that work was attacked for its lack of rigor by Core-Plus, but still merited attention, because the findings contradicted those of Core-Plus' own evaluation.
Unfortunately, the research of Bachelis and Milgram stood out, perhaps, more than it should have, because it stood alone. The reform movement had not matured enough to support similar studies. Such work has only recently become possible, as the first students involved in reform-based mathematics projects graduated from high school and entered the doors of colleges and universities.

Finally, much literature existed on the importance of college-related performance measures. This research supported the need for and use of ACT and SAT scores, freshman mathematics grades, and initial choices of majors in the study at hand. Especially, the work of Adelman (1998, 1999), based on data collected by the National Center for Educational Statistics, indicated a relationship between performance measures and degree attainment. McCormick (1999) correlated degree attainment to first year GPA, and Murtaugh, Burns, and Schuster (1999) established the same relationship for sophomore and senior years. Benefield (1996) established a strong correlation between ACT scores and successful completion of a pre-engineering program, and Levin and Wycoff (1990) determined those courses whose grades were predictors for persistence in engineering. Lewis (1996) and Kull (1996) conducted national descriptive studies, investigating remedial courses in colleges. Their work demonstrated the tendency toward remediation at the college level and some of the consequences of that phenomenon. This type of study was of interest here because remedial mathematics courses existed at Montana universities.

Why students choose their majors was important in this study, because initial choice of a SMET major is related to high school curricula and performance (Drew, 1996, p. 88). Dawson-Threat and Huba (1996, Strenta (1994), Drew (1996), Maple and
Stage (1991), and Pearson and Fechter (1994) examined the phenomena associated with choices of majors, and their work was considered herein.

Theme Development

Context of the Problem

**Problem Statement in Context.** It was not known if there were differences between former SIMMS mathematics students and those taught from more traditional high school mathematics curricula (non-SIMMS) in terms of ACT and SAT scores, freshman college mathematics grades, or initial declarations of college majors. Furthermore, a fundamental premise of this study was that the theoretical frameworks, goals, and expectations of the two program types, SIMMS vs. traditional, were different, and that difference confounded the process of comparing the achievements of students from both programs. Finally, the criteria against which achievement should be measured had evolved with the mathematics reform movement, resulting in the ambiguity of goals and outcomes for students. Despite the fact that traditional measures such as ACT and SAT scores and freshman course grades were important to this study, more modern measures were needed, based on professional standards, to round out a profile of mathematics performance for freshman students. Students' initial choice of majors and the survey of classroom practices served this purpose by adding reform-based criteria to the assessment.

The National Council of Teachers of Mathematics (NCTM) and the National Science Foundation (NSF) have been powerful forces behind the reform movement, each
serving the mathematics education community in a different manner. During the last
decade NSF has awarded millions of dollars to promote substantial change in the way
mathematics is taught, and its six Drivers of Reform (NSF, 1999) were meant to keep
awardees at the project and state levels on the reform trail. At the district, school, and
classroom levels, the NCTM Standards (NCTM 1989, 1991, 1995), three volumes of
them, were written to guide educators along that same path.

The evolutionary nature of the mathematics reform movement cannot be
overlooked, and the NCTM Standards and the NSF Drivers of Reform (National Science
Foundation, 2000) are more a product of that movement than a pre-movement handbook,
as this review will demonstrate. (See APPENDIX B for a concise listing of the Drivers of
Reform and the NCTM Principles and Standards.) Reform planners considered both the
Standards and the Drivers important, even though, ironically, neither was fully developed
until after large reform projects began. Consequently, those who planned and
implemented reform-based projects early in the movement lacked the guidance that the
refined documents could have provided.

Change From the perspective of educational change theory in the sense of Hall
and Hord (1987) or Fullan (1991), the mathematics reform movement was expected to
have evolved in phases. Those nine phases were described by Hall and Hord (1987), as
research, development, diffusion, dissemination, adoption, implementation,
institutionalization, refinement, and abandonment (p. 331). Evidence in the literature
affirmed that reform had evolved moderately in both secondary and post-secondary levels
Mervis (1998) and Green (1999) affirmed criticism of the mathematics reform projects, the rough spots along the reform trail, predicted by theorists, but there also were planned interventions to deal with pot holes (Hall & Hord, 1987, p. 142). It was not always possible to predict when or what interventions were needed, however, or to keep "a lot of plates spinning simultaneously in the air" (MCTM, 1997, p. 33) while in the process of implementing systemic change. Keeping all those plates spinning in the arena of statewide performance standards was a difficult trick in Montana, particularly, and that difficulty indicated a gap between dissemination and implementation phases of planned change.

More specifically, although Montana mathematics standards had been disseminated, (http://www.state.mt.us/usys/edu.htm), the state balked at implementing statewide assessment based on the tenets of reform, unlike some other states. Compare, for example, Vermont's progressive assessment system at the web site, http://www.FairTest.org. About Montana's current assessment practice, the assessment organization, FairTest, states "Montana has a bare-bones state assessment program that needs many major improvements. The state system relies entirely on multiple-choice, norm-referenced tests in three grades. This program should be replaced" (FairTest: The National Center for Fair and Open Testing, 2000, Montana data.). It is plausible, then, that Montana shared with some other states a more moderate phase of reform, as described in the change hierarchy of Hall and Hord (p. 331). More pointedly, Montana's educational institutions were not driven by a state-mandated reform agenda. Hence, there
were no statewide tests, based on the tenets of mathematics reform, with which to measure students' achievement.

Exemplifying a statement by Fullan, "Change is a Process, Not an Event" (p. 8), the guidelines for reform evolved with the movement itself. This evolutionary approach to reform was not without consequences, even if planned change was the reform strategy of the NCTM. One consequence was that the three volumes of "boiler-plate" Standards appeared and, indeed, evolved over a period of seven years. Likewise, NSF's Drivers for Reform, the guidelines meant to spur grant awardees down a narrower path of change, appeared mid-program for those awardees (MCTM, 1997, p. 1). The appearance of the NCTM Standards and NSF Drivers in this manner constituted rule-making with a play in progress. In any case, the latent appearance of the rules of the game undermined the belief that SIMMS, or any other reform-based curriculum project at that time, had access to all of the crystallized goals that would later appear in the NCTM Principles and Standards or the NSF Drivers.

Summarizing, because of the evolutionary nature of the reform movement, any equitable and valid comparison between the performance of SIMMS and non-SIMMS students was considered in light of the on-going changes. Appropriate measures were chosen as a compromise between reform and traditional philosophies in vogue at the time that students in this study took their mathematics in high school and college.

The Disparity Between Assessment Types. For entering freshman just out of high school, admission to Montana's two state universities required their ACT or SAT scores, high school grade point averages (GPA), and class ranks. (Montana State University
These measures, together with high school course grades and scores from several other standardized tests, traditionally became the trails of achievement left by graduating high school seniors. Although grade point averages and class ranks may have been program specific, standardized tests were not, since those tests were not designed to measure program specific knowledge, skills, or aptitudes. As Stake (1995) implied, even when a mathematics program's content closely matches that of some standardized test, that test fails to measure alternative constructs, specific knowledge, or useful skills (p. 173). Instead, standardized tests do purportedly measure those constructs agreeable to test writers, even when the test writers have little in common with those tested. Thus, a trail of standardized test scores may fail to indicate strong points of a mathematics program, instead measuring some selected concerns of disconnected "experts."

In contrast to traditional assessment practices, Romberg and Wilson (1995) argued for an authentic assessment framework based on the NCTM Standards to meet the challenge of a modern paradigm for mathematics education, founded, in part, on constructivism. Doing mathematics is neither a static nor a linear process. That process is too rich to be adequately assessed by standardized tests alone, despite pervasive traditional testing methods that assess static and linear knowledge. (Romberg & Wilson, 1995, p. 4) On the contrary, doing mathematics in a manner recommended by the NCTM Standards demands richer assessments:

If one considers mathematics to be a static, linearly ordered set of discrete facts, then the logical choice for a valid assessment system is the traditional standardized achievement test. On the other hand, if one views mathematics as a dynamic set of interconnected, humanly constructed ideas, then the assessment system must allow students to engage in rich
activities that include problem solving, reasoning, communications, and making connections. (p. 4)

The importance of the latter statement to this study could be denied: traditional and reform-based assessments clearly differed on theoretical grounds.

The Disparities in Program Assessment and Evaluation. "The primary question to be answered in any program evaluation is, How well is the program working in relation to goals and expectations for the students?" (NCTM, 1995, p. 66) Implicit in this question was the notion that program assessments should align with program goals and expectations for students. Assessment instruments that gather information about student achievement serve as focal points for a specific program, because they (ideally) measure those constructs that are valued within that program.

A central premise in this section was that values and practices differed among mathematics programs, confounding any attempt to measure the achievement of SIMMS and non-SIMMS students by traditional means alone. More specifically, at least four types of value-practice conflicts, described subsequently, arose in serving both groups, because the two types of curricula were so different. A second premise responded to the first by solving the problem implicit in its statement. That is, values and practices must align with assessment goals. Hence, for the purpose of this study, comparing the performance of former SIMMS and non-SIMMS students depended on the recommendations of secondary and post-secondary mathematics professional organizations and best-practice recommendations as well. Pertinent professional organizations included the National Council of Teachers of Mathematics (NCTM), the Mathematical Association of America (MAA), and the American Mathematical
Association of Two-Year Colleges (AMATYC). The well-known Seven Principles for Good Practice in Undergraduate Education (Chickering & Gamson, 1987) were also considered here.

At least four value-practice conflicts arose as a result of comparing SIMMS curriculum to more non-traditional curricula. Value-practice conflict one occurred from attempts to evaluate two essentially different programs, SIMMS and non-SIMMS, using only traditional standardized tests (Romberg & Wilson, 1995, p. 4). Notably, the state of Montana mandated standardized testing as a means of comparing programs. Montana's Office of Public Instruction (OPI) published its National Assessment of Educational Progress (NAEP), SAT, and ACT scores as state indicators of achievement and annually required districts to submit standardized assessment scores from grades four, eight, and eleven (Nielson, Lamson, Love, & Quinlan, 1999), thereby encouraging their use for program assessment. Furthermore, mandated assessment was becoming more rigid in Montana, as the state prepared to unify its assessment process. Consequently, districts lost the option to choose from several testing companies, as indicated on the OPI Web site (http://www.metnet.state.mt.us/MAIN.html, 2000). Many high schools, especially those in smaller Montana school districts, had little means to evaluate their programs otherwise, and they may have relied on state-mandated standardized tests to measure the quality or effectiveness of their academic programs. Therefore, what the NCTM standards espoused as important for program assessment, "a variety of high-quality evidence" (NCTM, 1995, p. 67), was more than the state of Montana required and, most likely, more than many schools could afford. The predominant use of standardized tests
certainly did not, however, provide the wide variety of indicators needed to evaluate programs or compare programs with each other.

The second value-practice conflict resulted from the predominance of traditional teaching methods in the college mathematics classroom, including the established use of summative assessments at the expense of formative assessments. LaBerge, Zollman, and Sons (1997) noted that although there were signs of change, "more traditional modes of instruction are still common in mathematics courses" (p. 15). Most likely, entering freshmen soon learned that testing defined college assessment, as they discovered that much of the personal attention associated with high school classes was not a matter of fact in college. In college classes homework might not have been be collected, and instructors might not have been available when students needed help. Course grades were probably based only on two or three tests and a final exam. Hence, teaching practices that were valued and recommended by the MAA, NCTM, or AMATYC may not have been valued or practiced in the freshman college classroom. Again, this system may have favored students coming from more traditional high school mathematics backgrounds, while putting former SIMMS students, who may have been adept at alternate modes of expressing knowledge, at a disadvantage. In the final analysis, course grades probably reflected traditional teaching methods, content, and processes. With this in mind, the Instructor Survey of this study provided a profile of classroom reform against which freshman performance might be better understood.

A third value-practice conflict involved the disparity between assessment of college mathematics programs and the assessment of student knowledge. This disparity favored tradition for two reasons: the faculty controls the curricula, and program
assessments are often lacking, even when departments desire reform. Each of these assertions deserves further examination.

First, the faculty tends to control the college curricula (Janzow, Hinni, & Johnson, 1997, p. 269). Also, "Most faculty members, whether or not they develop written objectives, tend to cast desired outcomes in terms of content acquisition rather than in terms of student outcomes" (Stark and Latucca; 1997, p. 269). Professional organizations support faculty control of curricula, too, as described in the recommendations of the MAA to its members, found in Guidelines for Programs and Departments in Undergraduate Mathematical Sciences (MAA, 1998). "The department's faculty should have the primary and most influential voice in setting the prerequisites, the course content, and the exit competencies for the department's courses" (p. 8).

Second, as a consequence of poor assessment management, programs may not incorporate feedback from evaluations to promote or reinforce change. Even when colleges assess their programs, feedback may never reach intended sources. Farmer and Napieralski (1997) support this notion, arguing that the lack of college program evaluations is not the problem. Rather, there may be more concern over doing the evaluations than with the use of findings, a focus which undermines the assessment process, encourages a cynical outlook among those involved, and reinforces "organizational inertia" (p. 602).

Consequently, faculty decide what is valued and practiced regarding student assessment. They tend to be traditional in their practices, and their attitudes may remain unaltered by superficial attempts at program assessment. As a result, what SIMMS students brought to the college arena, if not actively affirmed by departments, may never
have surfaced in college mathematics classrooms. Instead their abilities may have been sifted out of expression by assessments designed to measure only traditional content, constructs, and processes. Implied by this argument is that freshman course grades may have better reflected learning in a traditional way, rather than in a manner that is more consistent with reform practices.

A fourth value-practice conflict stemmed from what Artigue (1999) called "cultural gaps" (p. 1378) between high school and college mathematics programs, citing two theoretical frameworks to support his claim. According to the first framework, students assume a dualistic approach to learning. That is, they focus on beating the educational system as well as on learning course content. According to the second framework, the meaning of mathematical objects takes on an institutional flavor, that flavor, markedly different in secondary and college settings. Since much of our knowledge is contextual, Artigue claimed, building on what students bring to the classroom is important, and articulation becomes paramount to that process (Artigue, 1999, p. 1378). Hence, secondary and post-secondary programs would do well to exchange information.

The value-practice conflicts arising from cultural gaps between high school and college experiences may have been profound for college freshmen, especially while they learned the new (college) system. Applying these ideas to the study at hand, if college classroom practices did not conform to those of mathematics reform, the consequences for former SIMMS students may have been pronounced. Once again, the effects of the conflicts may have surfaced in course grades, putting former SIMMS students at a disadvantage in any comparison.
In summary, value-practice conflicts arose with differences in theories, content, practices, goals, assessments, between high school and college mathematics courses and programs, and even among courses and programs at the same educational levels. Specifically, they emanated from the inappropriate use of standardized tests for evaluating different mathematics programs, from the disproportionate use of traditional instructional methods in classrooms, from the disparity between program assessment and assessment of student knowledge, and from the cultural gaps between high school and college mathematics programs. Additionally, only the exceptional department, secondary or post-secondary, well-informs itself. Concerned by the lack of articulation between high school and college mathematics programs prompted the MAA to recently organize a Task Force to address the issues (MAA creates Task Force on Articulation, 2000).

The first premise of this section claimed that values and practices inherent in various mathematics programs differed, confounding the ability to compare the achievement of former SIMMS and non-SIMMS students. The four value-practice conflicts highlighted the issues associated with equitably and validly assessing SIMMS and non-SIMMS groups. Regardless of possible bias toward traditional content and processes, ACT and SAT scores were important comparative measures for the study at hand, as were freshman course grades. To mediate potential bias toward the traditional when comparing former SIMMS and non-SIMMS college mathematics grades, the survey of Classroom Practices and course Supervisor Interviews were warranted, since they help profile the extent of reform practice in the freshman mathematics classroom. In the next section, much will be said about professional standards to support the second
premise of this section, that a comparison of SIMMS and non-SIMMS students needed to be, in part, standards-based.

**The Need for Deep Change.** Both the NCTM and the NSF have encouraged a planned change approach to reform. Both organizations continue to expect, as outcomes of reform, what Fullan (1991) refers to as "second-order changes [which] seek to alter the fundamental ways in which organizations are put together, including new goals, structures, and roles" (p. 29). In their hope that program evaluation will facilitate reform, the authors of the *Curriculum and Evaluation Standards* (NCTM, 1989) referred to this notion as "deep structural change" (p. 237). Likewise, NSF's choice of the word, "systemic," in their State Systemic Initiative call for proposals boldly entitles the same objective: to encourage substantial structural change in mathematics education at the state level. (National Science Foundation, 1999) The drivers of reform clearly indicate that systemic means local and state-wide buy-in to the tenets of reform (National Science Foundation, 2000), and this type of encouragement goes hand-in-hand with planned change tactics espoused by Fullan (1991) and Hall and Hord (1987).

Whether fundamental change in mathematics education was taking place anywhere, particularly in Montana, depended in part on whether mathematics departments, secondary or post-secondary, followed the recommended assessment and evaluation guidelines found in various versions of professional standards. More strongly, the *Curriculum and Evaluation Standards* state the following.

Any program can be implemented by degrees. At one level, the language of the new program can be adopted while day-to-day instruction remains unchanged. At another level, minor changes in structure can be made by inserting a new unit into a course or slightly modifying the scope
and sequence. At a third level, deep structural changes can be made that include altering how people think about a program, how mathematics is presented, and how students come to know mathematics. The Standards speaks to this third level of change. It is the role of program evaluation to facilitate this deep change. (NCTM, 1989, p. 237)

In theory, then, if departments did not adhere to the guidelines for program evaluation, those deep structural changes called for by the NCTM Standards may not have taken place at the classroom level. It is at this grassroots level that the most meaningful outcomes of reform are measured by assessing student achievement. There can be no deep systemic change, if there is no deep classroom reform.

Feedback from the classroom is crucial to the success of any change process. That feedback must be built into assessment programs, according to evaluators and change experts (NCTM, 1995, p. 86; Fullan and Stiegelbaure, 1991, p. 183), but not merely because it provides comparative data. That feedback can also facilitate what Hall and Hord call "mutual adaptation" in deference to significant work by the Rand Corporation. As they further explain, "For successful change, the innovation [needs] to be adapted to the local setting, and the local setting [needs] to be adapted to match the innovation" (p. 331). For example, the evaluation standards (NCTM, 1999) encourage local control, recommending community representation on evaluation teams and the harvesting of data from student assessments that are oriented to local programs. (p. 247) Thus, as they measure effectiveness, program evaluations should also provide feedback for meaningful program interventions, tailored to the needs of local educational communities.

The Evolution of Standards and NSF's Drivers of Reform. Clearly, the evidence indicates that professional guidelines for reform evolved during the last decade with the
reform movement. The publication of the NCTM Curriculum and Evaluation Standards formally launched a period of change in 1989, but it would take six years for the organization to complete its three-volume set of guidelines. Although the core ideas appeared early in the first-to-press Curriculum and Evaluation Standards, those ideas were supplemented by the subsequent Standards for Teaching Mathematics (NCTM, 1991), and the Assessment Standards for School Mathematics (NCTM, 1995). Working drafts preceded all three of these volumes in an effort to involve the mathematics community in the mission and process of reform. Letters soliciting comments from the mathematics community precede the contents of each draft (NCTM 1987; NCTM, 1989; NCTM, 1993).

The first publication, NTCM's Curriculum and Evaluation Standards for School Mathematics followed a working draft in 1987. The framework therein is not overly prescriptive. Instead, the standards were broad enough to be accepted by a reticent but changing secondary mathematics community, yet narrow enough to suggest methods and goals for that community's embroiled transition toward reform. The Standards did not constitute a curriculum. Rather they were "value judgements based on a broad, coherent vision of schooling derived from several factors: societal goals, student goals, research on teaching and learning, and professional experience" (NCTM, 1989, p. 7).

Two years after its Curriculum and Evaluation Standards appeared, the NCTM published the Professional Standards for Teaching Mathematics (NCTM, 1991), which outlined the responsibilities in the process of reform not only for teachers, but for policymakers, schools and school systems, colleges and universities, and professional organizations as well. Among other instructional changes, those standards pressed for
teacher involvement in program reform and college faculty involvement in schools (p. 184). In 1995, three years after the NSF awarded its first multi-million dollar State Systemic Initiative grants, Assessment Standards for School Mathematics (NCTM, 1995) appeared in print. This volume addressed assessment more formally than did the Curriculum and Evaluation Standards, adding emphasis to several areas, including program evaluation. This document's final section, "Evaluating Programs" (p. 66), supplements the program evaluation standards in the Curriculum and Evaluation Standards, published six years prior.

In 1995, the American Mathematical Association of Two-Year Colleges (AMATYC) published its Crossroads in Mathematics: Standards for Introductory College Mathematics Before Calculus. The AMATYC standards, many of which mirror those found in the NCTM publications, have two major goals: "to improve mathematics education in two-year colleges and at the lower division of four-year colleges and universities and to encourage more students to study mathematics" (p. xii).

From 1990 to 1993 the Mathematical Association of America (MAA) developed its Guidelines for Programs and Departments in Undergraduate Mathematics, as stated on the MAA web page, http://www.maa.org/data/guidelines/guidelines%5Fintro.html. This set of guidelines, referred to as a "living document," was meant for revision and is now being revised. It has been distributed widely and was transmitted to other professional organizations as well, according to introduction at the web site. As a companion to the Guidelines and as a complement to the NCTM Standards at the university and college levels, the MAA published its Quantitative Reasoning for College Graduates: A Complement to the Standards. (MAA, 1998b) This document, developed
by committee, defined quantitative literacy for college graduates and set forth expectations for colleges and universities to meet their definition. Included in the document are specific goals, actions and strategies, and guidelines for assessment, many of which parallel the NCTM standards.

Note that the process of "standardization" has been examined formally for both secondary and post-secondary mathematics. Hiebert (1989) considered the relationship between educational research and the NCTM standards, and Robson and Latiolais (1999) looked at the implications of standards-based education for college faculty. Hiebert noted that, while research can influence the nature of standards, document the effectiveness of ideas, document the current educational situation, and suggest explanations for successes or failures, it cannot in general prove what is best practice. Too many variables confound the process (p.6). Just as importantly, research does not solely determine standards. Instead standards are "statements about priorities and goals. In education, they are value judgments about what our students should know and be able to do" (p. 5).

Robson and Latiolais sought to define standards-based education for post-secondary mathematics, pointing to the inevitability of this type of curriculum. (Robson & Latiolais, 1999, p. 8). While they did not discuss negative issues associated with standards-based learning, they did cite benefits. Adopting a standards-based framework forces discussion about educational issues, allows external reviews of student achievement, promotes clear communication with public and private stake holders, and enables better participation in the growing global market place (p. 4).
As with the evolution of the standards, the important "Drivers for Reform," listed by the NSF as core goals for State Systemic Initiatives, also evolved with the reform movement. Meant to prod the State Systemics down a grant-funded path of reform, these drivers crystallized NSF's original mission: to foster improvements in science, mathematics, engineering and technology (SMET) education "through comprehensive systemic changes in the education systems of the states" (National Science Foundation [NSF], 1999). (See APPENDIX B for the Drivers of Reform.) The latent appearance of these drivers potentially affected systemic operations at several levels and raises questions about SSI goals.

Despite the late appearance of the Drivers, those states receiving NSF funding developed specific goals for their grant proposals, even if what they produced was not quite what NSF eventually articulated. In the case of Montana's Systemic Initiative, NSF evidently informed the SIMMS project of the drivers in 1996, four years after the initial SSI awards, as indicated in SIMMS Monograph 3: Final Report (Montana Council of Teachers of Mathematics [MCTM], 1997, p. 1). Even so, the Montana-based SIMMS Project's accomplishments mostly agreed with the latent drivers, as their final report shows. More generally, however, because the drivers appeared well after Systemic Projects began, questions remain about any alignment of many of the Systemic Projects' goals and practices with the NSF Drivers of Reform. (Mervis, 1998)

A Consequence of Mutual Adaptation. Mutual adaptation, as described by Hall and Hord (1987, p. 116), can change the rules of the reform game, and, consequently, what is expected of players. While the evolution of the NCTM Standards has been both
solicited and visible, mutual adaptation in NSF policies may be less obvious. Of interest here, especially, is Driver 5, which changed from the time of its original dissemination. The pre-mutated form is of interest because it mandated the collection and analysis of SMET data. That requirement which, evidently, was relaxed mandated an indicator of the effect of reform-based high school curricula on students. The original Driver 5 was more specific in its direction than the updated version found at the NSF Systemic site as of February 22, 2000: http://www.ehr.nsf.gov/EHR/ESR/driver.asp. The original version, as stated in the Final Report of the SIMMS Project, prescribed the following.

Accumulation of a broad and deep array of evidence that the program is enhancing student achievement, through a set of indices that might include achievement test scores, higher level courses passed, college admission rates, college majors, Advanced Placement Tests taken, portfolio assessment, and ratings from summer employers, and that demonstrate students are generally achieving at a significantly higher level in science and mathematics. (MCTM, 1997, p. 2)

It should be noted that the newer rendition of this driver mandates reporting of achievement data by grade levels; it does not refer to the indices mentioned above. The discrepancies between the two versions qualify as evidence of the mutual adaptation, and it is clear that the tenure of the five-year grant awards made it difficult to collect all types of data suggested by the original Drivers.

The SIMMS Project began its work with ten objectives for systemic mathematics and science reform in Montana (MCTM, 1997, p. 1). In none of these original objectives can be found the promise to assess the effects of their high school curriculum after grade twelve. In objective two, however, the SIMMS Project promised to redesign the 9-12 curriculum, and in its promise is a tacit obligation to test that curriculum. The project did evaluate its curriculum as it was piloted and later conducted studies (MCTM, 1998) of
college freshman who had been SIMMS students. These studies, however, were limited to some descriptive data and interviews with the students (Allinger, Lott, & Lundin, 1998).

Clearly, then, documents meant to guide the reform movement through the phases of change in the sense of Hall and Hord (1987, p. 331) did not exist in the movement's infancy. Rather, they evolved with the movement, perhaps as a consequence of mutual adaptation. None of the NSF drivers appeared until four years after State Systemic awards. Specifically, the original NSF Driver 5 called for evaluation of reform-based mathematics programs that would extend beyond high school, but that driver also evolved with the movement. The SIMMS Project's original goals did not provide for post-secondary assessment of student achievement, although its staff chose to do so.

For the SIMMS Project and other State Systemic Initiatives that produced high school curricula, a clear NSF mandate to assess the achievement of college freshmen, if obeyed by the projects, might have constituted an planned intervention, so critical in the management of change. Importantly, however, the five-year length of the Systemic Initiative grants, unless they were renewed, provided too short a time-frame to adequately assess college performance subsequent to four years of curriculum development and pilot testing. Ironically, the latent or relaxed appearance of Driver 5, a symptom of the mutual adaptation process, may have undermined the amelioration of criticism of a novel curriculum.

The NCTM Standards themselves continued to weather severe criticism (Green, 1999; Mervis, 1998), despite initial hopes for adaptation within the mathematics community. It is conceivable that, if reform guidelines had been clear from the
beginning, planned interventions in the sense of Hall and Hord (1987, p. 141) might have thwarted some of the destructive criticism that plagued the reform movement. On the other hand, as change theory predicts, a fidelity approach to change, one that demands strict adherence to rules and disregards the concept of mutual adaptation, also begs for resistance.

Two Levers for Reform: Curriculum and Program Evaluation. The use of the written materials to leverage reform was a key strategy in the NSF reform plan, and that strategy buttresses a critical premise for this study. That is, the SIMMS curricular materials were substantially different from other materials in use during the treatment period, 1994-1998. The assumption that written materials could lead the way to reform was operationalized by NSF and by the SIMMS Project, as evinced by respective drivers and objectives. With utter conviction, NSF Driver 1 called for "implementation of comprehensive, standards-based curricula as represented in instructional practice, including student assessment, in every classroom, laboratory, and other learning experience provided through the system and its partners" (MCTM, 1997, p. 1; NSF, 2000). Similarly, curricular implementation was a major goal of the SIMMS Project, becoming formally embodied as the project's Objective 2: "Redesign the 9-12 mathematics curriculum using an integrated interdisciplinary approach for all students" (MCTM, 1993, p. 1; MCTM, 1997, p. 6). While lamented by some educators, the documented overuse of texts in the classroom, especially by mathematics teachers, turned out to be a primary strategy for reform, used by NSF and the reform-based projects
themselves. Solid evidence supporting this practice comes from the meta-analysis of research by Robitaille and Travers (1992).

Teachers of mathematics in all countries rely very heavily on textbooks in their day-to-day teaching, and this is perhaps more characteristic of the teaching of mathematics than of any other subject in the curriculum. Teachers decide what to teach, how to teach it, and what sorts of exercises to assign to their students largely on the basis of what is contained in the textbook authorized for their course. (p. 706)

In deference to this finding, the study at hand utilizes as fact that curriculum materials lead the way to reform. Indeed, a major goal of NSF funding of the State Systemic Initiatives was to saturate the educational arenas with reform-based curricula to begin the march toward reform (NSF, 2000).

The question, then, becomes one of deep change. While the curriculum can guide initial progress toward mathematics reform, can it alone provide the depth sought by reformers? Lessons learned from previous reform efforts such as the “new math” movements in the 1960s demonstrated that more than materials are needed to produce a desired effect (Kilpatrick, 1992, p. 24). When observing teachers, experiences such as those of Dapples (1998, p. 78) indicate that materials alone may not be enough to coax teachers to change the way they relate to students, despite the fact that they make surface-level adjustments to their practices. In general, non-traditional content can be taught in a most traditional manner, as the Standards themselves indicate (NCTM, 1989, p. 237), thereby leaving untouched the deeper issues associated with classroom practice. From a change theory perspective, such behavior merely constitutes one phase of the change process thought to progress in stages (Hall and Hord, 1987, p. 330). However, if program evaluations emphasize methods of teaching and assessment that align with standards, in
addition to supporting a Standards-based curriculum, avoidance of critical classroom
reform issues may be more difficult. Accordingly, well-structured and executed program
evaluations can facilitate the change process by guiding teachers through the phases of
change. Again, it is the notion of deep change, considered previously, that is thought to
be a consequence of solid program evaluations, a second powerful lever with which to
force grassroots reform.

Thus, it is thought that curriculum materials first lead the way to reform and
program evaluation builds the momentum toward deep change. Importantly and
conversely, if program evaluation does not favor curriculum materials that embody
standards, there may be little hope that the respective program agrees with them. The
extent to which high schools and post-secondary mathematics departments align their
programs to professional standards is beyond the scope of this study, but gauging this
alignment is still extremely important in assessing the effectiveness of the reform
movement.

Within the scope of this study, though, is the measurement of classroom practices
of college mathematics instructors and how these align with professional standards. The
results of the instructor survey at two Montana universities were compared with those of
other studies to provide a backdrop for conclusions about comparative performance of
Montana university freshmen in their mathematics classes. Additionally freshman
mathematics course supervisors were interviewed about the goals, expectations,
standards, and outcomes they have for the students in their classes.
The STEP Project: A Partner in Reform. The National Science Foundation did provide substantial funds to other projects that promised to reform mathematics and science teaching at the college and university levels. The Systemic Teacher Excellence Preparation (STEP) Project was such a program in Montana that affected mathematics teaching at the secondary and post-secondary levels. STEP began operations in 1993 as a six-million dollar, five-year project with a mission to reform K-16 mathematics and science teaching in the state. Primary funding for STEP ended in 1998, with the annual report listing five lower division mathematics courses at UM and four at MSU benefiting from reform-based revision along with a host of other science and mathematics courses also affected. Overall, the revisions in these courses included the adoption of reform-based texts, incorporation of technology and cooperative learning into classrooms, the use of problem-solving approach to teaching, the use of manipulatives in classrooms, and the incorporation of writing for reflection. Different revisions were implemented in various courses (Systemic Teacher Excellence Preparation [STEP], 1998, pp. 124-129).

It should be mentioned that the STEP Project was multifaceted and that post-secondary course reform was one goal of several. The STEP Year Five Annual Report (STEP, 1998, p. 124) clearly indicated the specific revisions made in each course, although there was no mention of a survey of classroom practice, such as the one proposed in this study. Nevertheless, as a result of STEP efforts, Luebeck (1998) and Thoreson (1997) both noted substantial grassroots involvement in the preparation of new teachers. That preparation process itself was steeped in the tenets of mathematics and science reform.
Current Understanding of the Problem

Modern Standards for Mathematics Education. NCTM's recently published Principles and Standards for School Mathematics (NCTM, 2000) represents a condensation and revision of the three original NCTM documents governing, respectively, curriculum and evaluation, pedagogy, and assessment. The content of the Principles and Standards splits into three categories: principles, content standards, and process standards. The spirit of the Principles and Standards suggests an integrated approach to teaching and learning mathematics, mixing content as well as principles and standards into all curricula, as the introduction to the new document explains. (MCTM, 1998). However, even if content is separated by subject matter, as is traditional, both original and revised versions of this set of the NCTM guidelines suggest that principles and standards be incorporated across mathematics curricula.

The American Mathematical Association of Two-Year Colleges (AMATYC) borrowed much from the original NCTM Standards. The AMATYC document is a set of guidelines that are "consistent with frameworks presented in other mathematics initiatives and [these] are intended to affect every aspect of introductory college mathematics" (AMATYC, 1995, p. 9). The document also splits guidelines into three types: Standards for Intellectual Development, Standards for Content, and Standards for Pedagogy.

The Mathematical Association of America's Guidelines for Programs and Departments in Undergraduate Mathematical Sciences is less prescriptive in the areas common to the previous two resources. However, one section entitled, "Curriculum and Teaching," reads much like the NCTM and AMATYC documents. Here the MAA
prescribes "quantitative literacy" in the following manner, under the subheading, "Mathematics and General Education."

Colleges and universities should strive to ensure that every graduate has achieved quantitative literacy in the sense of being able confidently to analyze, discuss, and use quantitative information; to develop a reasonable level of facility in mathematical problem solving; to understand connections between mathematics and other disciplines; and to use these skills as an adequate base for lifelong learning. (MAA, 1998, p. 8)

Other recommendations for college departments in this section of the MAA Guidelines include using technology, making connections to other disciplines and their content, and making use of research on learning and teaching.

Written after the Guidelines, MAA's Quantitative Reasoning for College Graduates: A Complement to the Standards (NCTM, 1998b) develops the notion of quantitative literacy, noted previously. As in the NCTM and AMATYC documents, Quantitative Reasoning specifies similar process goals for students, including graphical, symbolic, numerical, and verbal representation and communication of information. (Part II, p. 3). In Quantitative Reasoning is found a realistic assessment procedure (Part IV) with a recommendation for "mathematics across the curriculum" (Part III, p. 2). The document specifies methods of teaching quantitative literacy, including facilitative teaching, using collaborative learning, and integrating real-world problems into the curriculum. (Part III, p. 6). Additionally, in Appendix C the authors point to successful college and university reformed programs. In light of these recommendations, the authors purposefully reiterate the importance of teaching remedial courses "subversively," as argued in the dated, but vital paper, "Minimal Mathematical Competencies for College Students." (Committee on Undergraduate Programs in Mathematics, 1982).
"Students entering college with mathematical deficiencies have presumably had opportunities to learn the mathematics, and for them those opportunities did no work. Therefore, the college remedial course should not be a mere rehash, and certainly not an accelerated one, of the traditional secondary or even elementary course. Courses that cover the same old ground in much the same old way tend to be just as uninspiring and unintelligible for these students as the originals, and therefore even less likely to succeed. Students should be able to find even remedial courses fresh, interesting, and significant." (Part III, p. 5)

Hence, many of the ideas found in the NCTM and AMATYC volumes were iterated in the two MAA documents as well. Notably, the newly synthesized Principles and Standards for School Mathematics (NCTM, 2000) recapitulates and affirms the importance of those guidelines in the prototypical NCTM, MAA, and AMATYC documents. For two reasons this was important to the work at hand. First, the SIMMS curriculum is founded on prototype standards; the release of the newer Principles and Standards affirms their solid grounding. Second, any measurement of the application of standards, such as in the Instructor Survey in this study, is affirmed by the new document, given that professional guidelines at secondary and post-secondary levels converge in their recommendations for mathematics education.

The Core-Plus Project: Research Meets Practice. The United States has no national curriculum. However, in 1994 Congress directed the Office of Educational Research and Improvement (OERI) to evaluate curricula and to report exemplary and promising programs (United States Department of Education [USDE], 1999). As a result of this directive, an expert panel chose five exemplary and five promising mathematics programs, announcing their findings to the public in 1999. All of the chosen projects had
roots in the NCTM Standards, including the exemplary Core-Plus Project based at Western Michigan University.

After reports of students' college mathematics placement woes triggered controversy in the West Bloomfield and Bloomfield Hills School districts of Michigan, resident and Wayne State University mathematician, Gregory Bachelis, conducted a survey of 1997 graduates at two local high schools. His research consisted of comparing the attitudes and college related performance of former Core-Plus students with that of non-Core-Plus students. All data was self-reported by former students (Bachelis, 1998, p. 2). Bachelis himself did not evaluate much of the data except for initial responses to his questionnaire. Instead, he sent the data to R. James Milgram, a Stanford mathematics professor, for analysis (Milgram, 1999, p. 1).

Milgram's analysis of the Bachelis data put the Core-Plus curriculum at a disadvantage, if students' self-reports are to be believed. Mean mathematics ACT and SAT scores for the Core-Plus group fell below those of the traditional group (p. 5). Proportionately more former Core-Plus students took remedial mathematics courses during their first year at college than did the non-Core-Plus students (p. 3). Mean scores for first-year college mathematics classes also favored the non-Core-Plus group (p. 5), and students' comments were especially critical of the lack of algebraic content in the Core-Plus materials (p. 1). It should be emphasized that the Bachelis-Milgram work has been criticized by Core-Plus for its lack of rigor (Core-Plus, 1999), and the Core-Plus Project later produced a comprehensive, but glossy, evaluation of its efforts (Core-Plus, 2000b). Predictably, the results of that evaluation contradict the findings of Bachelis and Milgram.
Eventually, in a subsequent posting to his web site, http://www.math.wayne.edu/~greg/AppendI.htm, Bachelis placed results from University of Michigan studies and results from a Core-Plus hosting school, which tend to contradict his and Milgram's work. Clarity here may come only with more careful evaluations of that project, particularly because self-interest seems to be at the root of both studies in the cluster, and because research methods were less than rigorous.

The SIMMS Curriculum and College Freshmen. To gain reform leverage, the SIMMS staff operationalized research in several areas: textbook use, constructivist theory, technology use, and problem-solving in mathematics education. As stated previously, research indicated that teachers substantially rely on textbooks (Robitaille & Travers, 1992, p. 706), making the curriculum a wise choice for promoting mathematics reform. The nature and structure of the SIMMS curriculum materials were based in part on constructivist theory. Additionally, the materials extensively incorporated problem-solving and mathematical modeling of real-world phenomena. The curriculum strongly encouraged technology for building concepts as well as for calculating. Along with problem-solving and the use of technology, cooperative learning was integrated into most of the curricular activities. (MCTM, 1996-1998) The curriculum was written in response to a need and a national call for a mathematics program with the aforementioned qualities, as is crisply stated in the SIMMS Monograph 1: Philosophies.

In order to accomplish a shift in paradigms, the SIMMS Project recognizes the need for an entirely new curriculum—one that embodies the constructivist notions of doing, learning, teaching, and assessing mathematics. The decision follows the recommendations of the [NCTM] Standards' "Next Steps" (p. 252). (MCTM, 1993, p. 6)
Evidently, the SIMMS materials were novel and unique at the time of their use by the
student subjects of this study. However, other NSF sponsored products, such as Core-
Plus, eventually competed with SIMMS in an emerging market of materials purportedly
based on the NCTM Standards.

After several years of reform programs scaling up, the first graduates began
entering colleges and universities in the middle of the last decade. In the case of the
SIMMS Project, research on college-related criteria had to wait until enough student
veterans of the SIMMS curriculum entered the post-secondary institutions.

Allinger, Lott, and Lundin (1998) conducted interviews of the first two cohorts of
former SIMMS students entering college and gathered performance data from six
Montana colleges and universities, including The University of Montana and Montana
State University. Although the number of former SIMMS students was limited (n=84),
data from interviews and performance in mathematics classes indicated that the SIMMS
group differed from the general population of traditional freshman in their attitudes and
performance. While the proportion of former SIMMS students who took mathematics
courses in Fall 1997 exactly matched that of the traditional population, their performance
did not. Compared with the general traditional freshman population, the SIMMS group
placed into lower developmental courses proportionately more often and into mid-level
courses less often, although placement into calculus classes was approximately the same
as that of the traditional freshman population. Also, those SIMMS students who placed
into the developmental courses had a lower pass rate than that of the general population,
but that rates were similar for the calculus courses. Contrarily, those who placed into the
group of courses above developmental but below calculus did proportionately better than the general traditional population of freshmen (p. 17).

Interviews with students supported the outcomes of this performance analysis. Positive comments about SIMMS tended to revolve around students feeling prepared for courses like statistics and matrix algebra, both middle level freshman courses. Positive comments about standards-based "process" activities such as problem-solving, cooperative learning, using technology, and writing were recorded. Negative comments heavily pointed to the lack of preparation in traditional algebra and related material that dominates developmental classes and permeates traditional calculus courses (p. 19). The researchers concluded that many former SIMMS students felt caught between educational paradigms (p. 29). What was emphasized in their high school mathematics classes seemed to be different from what was emphasized in their college classes.

The Current State of Standards-based Practices. The work of Weiss, Upton, and Nelson (1992) and Weiss, Matti, and Smith (1994) were major efforts to measure the classroom practices of teachers as they align with the NCTM Standards. The first work, in the authors' own words, was designed to be "exploratory in nature," (p. 1) but from this pilot study came the prototype instrument used later by the second group of researchers in an extensive nationwide survey of teachers beliefs and practices. Although these two research articles are now dated, both represent snapshots of beliefs and practices of classroom teachers several years after the publication of the first NCTM document.

More to the point, LaBerge, Zollman, and Sons (1997) adapted the Classroom Practices Inventory (CPI) of Weiss, Upton, and Nelson (1992) for their interviews of 30
mathematics faculty in eleven Midwestern colleges and universities. This group of researchers reported that mathematics faculty, while still heavily favoring traditional lecture style delivery of content, seemed to be trying new methods. Also, as a group, they were more aware of the professional recommendations directly aimed at the college and university level than they were of the NCTM Standards (p. 16). The CPI used by LaBerge, Zollman, and Sons will be employed in this study without modification. (See APPENDIX A for the CPI.)

From the dissertations of Kull (1996) and Hopkins (1998) came other measurements of reform at the college level. Kull examined awareness of the NCTM standards and standards-based practice in independent remedial college and university programs. His survey of programs indicated that "there were some Standards-type behaviors going on in the developmental mathematics programs surveyed" (p. v). Hopkins (1998) studied the effects of standards-based high school mathematics on the self-concepts and achievement of college freshmen. Her findings suggest direct effects of Standards-based programs on both achievement, as measured by the ACT mathematics test, and on students' self-concepts, as measured by survey responses (p. 102).

College-related Measures, Performance, and Graduation. For the work at hand, three types of college-related measures were necessary: ACT and SAT scores, freshman mathematics grades, and students initial choice of majors. The two state universities of interest in this study required either ACT or SAT scores of incoming freshman (Montana State University, 1999; The University of Montana, 2000), and Montana State University used scores from either exam to guide placement into courses (Montana State University-
Bozeman, 2000). While scores should not be interchanged because their content differs (Jeff Schiel, Ph.D., Senior Research Associate for ACT, personal communication, January 28, 1999), the high correlation (.92) between ACT I and SAT scores enables such use, especially since psychometric properties are not an issue. (Dorans, N., Felicia, L., Pommerich, M., & Houston, W., 1997, p. 30) ACT prepares packets of conversion tables for the purpose of analysis (ACT, 1998), based on data from 103, 525 students at 14 universities (Dorans, N., Felicia, L., Pommerich, M., & Houston, W., 1997, p. 26)

Both ACT, Inc., and the College Entrance Examination Board, creator of the SAT I exam, established their tests as predictors of freshman grades (ACT, Inc., 2000; College Entrance Examination Board, 2000), while McCormick (1999) completed the link between freshman performance and graduation, finding the latter correlating highly with the former. Other data confirming the strong links among admissions tests, freshman performance, and graduation from college came from Adelman (1998, 2000) Murtaugh, Burns, and Schuster (1999), Benefield (1996), and Levin and Wycoff (1990). Their contributions were considered in more detail in the next section. It is clear, in summary, however, that ACT and SAT scores and freshman mathematics grades were measures that are linked to graduation. For the purposes of this study, they were credible measures with which to distinguish differences in performance between former SIMMS and non-SIMMS students.

The final college-related measure of interest in this work, was initial choice of major. It was important to know if former SIMMS and non-SIMMS students differed in their choices, especially with respect to SMET majors. Dawson-Threat and Huba (1996) referred to Chickering's monumental psychosocial theory of student development,
suggesting that thoughtful declaration of a major constitutes "clarifying purpose" (p. 297). That is, choosing a major may itself indicate development. Vetter (1994) argued that the basis of major choice may begin in childhood, and she cited ample evidence that students drop out of mathematics and science in large proportions before they reach college. Strenta (1994) in a large study of engineering students correlated and ranked achievement indicators that predict interest in science courses. He demonstrated that the best predictors of interest were high school mathematics and science grades, followed by the number of those courses taken. While major choice was clearly a complex issue, a foundation in psychosocial theory combined with emerging models of student interest and their early choices pointed to the importance of comparing majors among former SIMMS and non-SIMMS students. Clearly, initial major choice was an indicator of interest, preparation, and, possibly, psychosocial development. The research was considered more deeply in the following section.

Review of Previous Research

NCTM Standards-based Instruction and College Freshmen. Hawkins (1998) sought to examine causal relationships between high school mathematics instructors and students' mathematics self-concepts using two instruments, the Mathematics Aptitude Inventory and the ACT Mathematics Usage Test. The first instrument measured student perception of mathematics teachers, including their alignment with the NCTM Standards. The second instrument measured mathematical achievement, especially problem-solving ability. The researcher reported solid validity and reliability data for both instruments (p. 45), and she used a structural equation or "LISREL-type" (p. 47) modeling process with
the SAS procedure, CALIS, to analyze the structural relations in the theoretical model. With this procedure, the researcher was able to ascertain whether the model constructs were reliable and valid and then predict causal relationships among certain variables. Her result was a "structural model" of causal relationships on student self-concept (p. 48). Hawkins used cluster sampling to choose representative groups from 1997 Oklahoma State University colleges for a total sample size of \( n=331 \), or ten percent of the population, and she determined that the sample represented the population with respect to ACT scores. The researcher was able to determine a significant combined indirect effect of \( .30 \) (\( \alpha=.05 \)) due to NCTM Standards utilization on mathematics self-concept and effects of \( .28 \) and \( .41 \) on interest in mathematics and mathematics achievement, respectively. Notably, the total effect of the NCTM Standards on self-concept was higher than that of achievement (.21). Also, the direct effect of Standards utilization on achievement was .55, compared with lesser effect of the number of mathematics classes taken (.34). One other effect is important because of the implications for the practices of college instructors. Perceptions of mathematics teachers had direct effects on achievement (.39) and perceived usefulness of mathematics (.27).

Two Contradictory Core-Plus Studies. A review of the Core-Plus study by Bachelis (1998; 1999) and Milgram (1999) was warranted, because, despite its weaknesses, it exposed the college-related measures considered in this study. Since the Core-Plus Project was a reform-based project, the Bachelis-Milgram work also teased out some of the consequences of considering a new curriculum in light of old evaluation
methods. Finally, this work was a reminder that, if the design of the study suffers, so will its credibility.

Following a controversy surrounding lower than expected scores on mathematics placement exams at the University of Michigan, Bachelis surveyed two groups of 1997 high school graduates from separate but demographically comparable Michigan high schools. Andover High School had adopted the Core-Plus curriculum as it continued to phase out its more traditional math program, and Lasher High School remained traditional. Bachelis elected to compare the Core-Plus cohort at Andover, which included none of the accelerated group of students at that school, with what he considered to be a similar group at Lasher, all of those students who did not take calculus during their final year there. (Note that the validity of this match is questionable, since an unstated number of students in the Lasher group had been in Lasher's accelerated program but opted out of calculus.) To compare responses, Bachelis also decided to survey the accelerated students at both high schools, so that total populations at each school had the opportunity to respond.

Bachelis received 112 of 228 (49%) replies from former Andover Students, and 75 out of 251 (30%) replies from former Lasher students. It should be emphasized that all data were self-reported by students. Bachelis analyzed responses to two questions that generated descriptive statistics, and he also analyzed the written responses to questions concerning the effectiveness of the respective high school mathematics programs. Bachelis' preliminary analysis of the data points strongly to former Core-Plus students' dissatisfaction with their high school mathematics program in light of the different expectations inherent in their college courses. Of particular concern to these students was
a lack of preparation for the more traditional algebraic content encountered in college mathematics classes. It must be noted that written comments by students in the Lasher (non-Core-Plus) group were also very negative, disparaging either their high school classes for lack of preparation or their college classes for not being up to their expectations. However, comments from accelerated groups were more positive, connoting better experiences with mathematics in high school and college courses.

Bachelis sent his data to Milgram (1999) for more detailed analysis. The results of Milgram's analysis are best stated in his words.

The data represents a first glimpse—to the best of our knowledge—of how students trained in this new way perform in a university environment, and frankly, the results are not encouraging. First, almost all the Andover students were severely critical of the program, many bluntly blaming it for their difficulties in university level courses. Moreover, there was no measure represented in the survey, such as ACT scores, SAT Math scores, grades in college math courses, level of college math courses attempted, where the Andover Core Plus students even met, let alone surpassed the comparison group of Lasher students. This held true even when the group of Lasher students was restricted to those who had not even taken a pre-calculus course in high school. (p. 1)

In his analysis Milgram noted that the average mathematics SAT and ACT scores for the Core-Plus groups were 531 and 23.35 respectively, while the averages for the non-Core-Plus group were 590 and 25.09. The researcher also noted that average high school GPAs were similar for the two groups, 3.23 for the Core-Plus group (n=67) and 3.29 for the non-Core-Plus group (n=41). A $\chi^2$ test indicated that significantly fewer Core-Plus students placed into calculus and significantly more, into courses below calculus, when compared with the non-Core-Plus Students(p. 3). Milgram disaggregated the "remedial" (before calculus) category, and found significance favoring the non-Core-Plus students, but this disaggregation didn't add much to the analysis, since large differences remained
centered within calculus and remedial classes. Milgram's analysis of the freshman course grades involved the means for the Core-Plus and non-Core-Plus Groups. The mean Core-Plus math GPA was 1.9, while the non-Core-Plus mean was 2.6. Milgram's analysis of student comments reiterates the work of Bachelis.

It was easy to criticize the methodology of Bachelis and Milgram. Their sampling method, self-selection, was not rigorous enough to make inferences back to the populations of interest. In addition the data was self-reported. The Core-Plus Project noted these weaknesses, and formally rebutted the Bachelis and Milgram study on their web site (http://www.wmich.edu/cpmp/). Core-Plus also noted that Andover had been using pilot materials during the time of the Bachelis study, and that those materials could not represent the finished product. The Core-Plus rebuttal referred to research done by the University of Michigan registrar, noting that "graduates of the Core-Plus program perform as well as, or better than, graduates of a traditional mathematics program." (Core-Plus, 2000)

Interestingly, to rebut the Core-Plus rebuttal, Bachelis published the University of Michigan Registrar's data in an Update of his findings (Bachelis, 1999). As Bachelis himself noted (p. 7), the registrar failed to disaggregate the data into Core-Plus and non-Core-Plus groups, focusing instead on school-based means in calculus courses and on courses below calculus. Schoolwise, the average mathematics grades of former Andover students compared very favorably with those of former Lasher students. This data was descriptive, however, and no tests for significance were reported.

Also of note, the chairperson of the Andover High School Mathematics Department, in response to the Bachelis study, found that the number of Andover
students placing into calculus in 1997 jumped by 11% over the previous year at the University of Michigan (Bachelis, 1999, Appendix A). This was the same year of Bachelis' survey, and the evidence weighed against his findings. However, this data was not disaggregated into Core-Plus and non-Core-plus groups, nor were tests for significance done. Finally, Milgram noted that Andover's mean school ACT math scores showed a comparative decline since Andover's adoption of Core-Plus. (Milgram, 1999, p. 25) However, again, there were contradictions, including the highest gain in mean score in the first year that Core-Plus students would have taken the ACT.

The work of Bachelis and Milgram was anything but conclusive. The design of the study was weak, relying on self-reported data and self-selected sampling. Validity was questionable if, during the "treatment" of the Andover students, curriculum materials were in pilot form. Nevertheless, it is difficult to deny that some phenomenon was working within this turbulent system in Michigan. Evidently this phenomenon had its roots in change and in the tension generated by new expectations of the reform movement set against values and practices of the traditional mathematics system. Unfortunately, the lack of structure in the evaluations just reviewed critically wounded any credibility in the conclusions of these researchers. Any further study must incorporate stronger design to avoid the weaknesses evident in the aforementioned research.

For all of its shortcomings, the Core-Plus study did flush out important variables. While ACT and SAT scores represent more traditional measures of achievement, they are important mostly because they open college doors, or close them. Just as important, freshman grades represent achievement during a college year of highest cognitive gains (Pascarella and Terenzini, 1991, p. 160), and should reflect readiness to achieve at even
higher levels. Both measures were critical to the study at hand, and they deserved to be examined again in a more exacting manner.

The Core-Plus Project published its own in-house evaluation, Evaluation Results, which included college-related measures comparing Core-Plus and non-Core-Plus students. The glossy evaluation was meant for public consumption, having been printed by Core-Plus publisher, Everyday Learning. (Core-Plus, 2000b) Not surprisingly, the findings in Evaluation Results refute those of Bachelis and Milgram. Evaluators reported a non-significant (α=0.05) higher mean SAT mathematics score of 552.0 for Core-Plus Students (n=371) compared with non-Core Plus (n=190) mean of 543.4. The subjects for this study came from eleven schools (p. 7). On the ACT mathematics test, non-Core-Plus students (n=111) scored higher than the Core-Plus group (n=531), with mean scores of 19.8 and 19.2, respectively. P-values were not stated for either test. Evaluators noted that the science reasoning scores may better reflect the ability of Core-Plus students than the mathematics scores, and Core-Plus students' mean score was higher than their traditional cohorts, but not significantly so. No p-values were given for either SAT or ACT scores.

Core-Plus evaluators compared placement test scores among Core-Plus (n=164) and non-Core-Plus (n=177) students at one major university that used an MAA placement exam with subtests in basic algebra, advanced algebra, and calculus readiness. Core-Plus students scored significantly higher than non-Core-Plus students in calculus readiness with mean scores of 12.9 and 10.5 respectively, but there were no significant differences in the other two categories. No p-values or α-values were stated. The evaluators also note that data from the mathematics department indicated proportionately more students from the Core-Plus group placed into calculus, and fewer into precalculus
and intermediate algebra. At the developmental level proportionately more Core-Plus
students placed into beginning algebra. No results for analysis of significance were
provided (p. 9).

In the Core-Plus report, the results on college mathematics performance were
confounded by several factors. First, data came from students completing the pilot Core-
Plus program rather than a field-tested version. Second, the character of the mathematics
program at the pilot schools changed during the time that trend data was collected (Core-
Plus, 2000b, p. 10). Additionally, although Core-Plus evaluators concluded that "This
preliminary school-trend evidence suggests that students who experienced a pilot CPMP
curriculum were at least as well-prepared for calculus (AP or college level) as students in
more traditional curricula," no p-values or levels of significance are stated. Also, other
than mean math GPA, only data for courses at or above precalculus were provided. It
must be noted that students in this study came from two schools in a "suburb with many
affluent, well-educated residents" (p. 10). Thus, results could not be inferred to a more
demographically diverse setting.

The SIMMS Evaluation of Student Outcomes. Allinger, Lott, and Lundin (1998),
conducted interviews of incoming SIMMS Freshman in Spring 1997 and again in Spring
1998 at two Montana state universities. In addition to the interviews the researchers
collected data on the performance of college and university freshmen for Fall Semester
1997 from six Montana institutions of higher learning, including the two state
universities. All SAT, ACT, and course grade data came from institutional records, and
no self-reporting, other than from the interviews themselves, occurred in the data collection process.

Entering the six institutions in Fall 1997 were 4880 traditional freshmen, of whom 21% did not take mathematics courses that semester, 32% enrolled in developmental courses, 33% enrolled in mathematics before calculus, and 18% enrolled in calculus classes. Of the 4880 traditional entering freshman, 88 qualified as having three or more years of SIMMS mathematics. 79% (n=81) of those took mathematics classes that fall, which exactly matched the percentage of the general population of traditional freshmen taking mathematics. (Allinger, Lott, & Lundin, 1998, p. 16)

From those traditional freshmen students who took mathematics courses Fall 1997 the following comparisons were noted. A disproportionate number of SIMMS students enrolled in developmental courses, 56% (n=37), compared with 41% of the traditional freshmen. In mathematics below calculus courses, the percentages were again disproportionate. 27% (n=18) of the former SIMMS students competed with 41% of the general population. In calculus classes, 17% (n=11) former SIMMS students enrolled compared with 18% of the traditional freshman mathematics population.

The researchers also compared pass rates for the SIMMS group against those of the traditional freshman mathematics group. 49% (18 of 37) of the former SIMMS students passed their developmental courses, compared with 60% of all traditional entering freshmen who took the courses. For courses before calculus, 94% (17 of 18) of the former SIMMS students passed, versus 84% of the traditional freshman population. In calculus courses, 65% (8 of 11) of the former SIMMS students passed compared with 74% of the entering freshmen.
As with the Core-Plus studies, freshmen who were interviewed expressed concern over preparedness for college courses, especially where algebra comprised a solid portion of the course content. Allinger, Lott, and Lundin concluded that 73% of former SIMMS students thought themselves well-prepared in those process areas recommended by the NCTM Standards, but that 48% would recommend a combination of traditional and SIMMS-like training to a sibling who would go on to college. Also noteworthy, some former SIMMS students thought that the placement test, which focused mainly on algebraic skills, was a barrier, forcing them to take developmental courses. In a synthesis of findings, the three researchers conclude that students wrestle with two disparate and competing educational paradigms.

These seemingly contradictory percentages imply that students find themselves in a quandary. How can they embrace the notion of mathematics as presented in an NCTM Standards-based curriculum, and at the same time contend with their image of traditional mathematics and college mathematics as narrowly focused on algebraic rules and procedures? (p. 29)

This research reiterated some of the conclusions claimed by Bachelis and Milgram, but the design of the SIMMS study is more sound. It is important to note that former SIMMS students felt under-prepared for algebra-based college courses. As in the Core-Plus research, veterans of the traditional curriculum placed more often into non-developmental courses. Placement into calculus occurs was about the same rate for each group, with former SIMMS students not performing as well as the general population. Conversely, former SIMMS students placing into courses before calculus did better than the general population.
Neither the SIMMS research nor the Core-Plus research attempts hypothesis testing, except for Milgram's $\chi^2$ test on course placement, which suffers from a lack of credibility due to self-reporting and self-selective sampling. Still, taken together, these two studies point to a need for more information on the effectiveness of a reform-based curriculum, and for information grounded in sound methodology. Accordingly, the results of the study at hand should tell if former SIMMS students and their more traditionally educated peers significantly differ in mathematics scores on the SAT and ACT tests, in mathematics grades disaggregated according to course type, and in their choices of majors. The methodology of this research promises to be stronger than previous studies of the Core-Plus Project or the SIMMS Project, although it will seek measures common to both.

The Meaningfulness of College-related Measures. Adelman (1999) conducted an extensive analysis of the relationships between resources students bring to college and their degree completion rates. Using data from the National Center for Educational Statistics (NCES) study, High School & Beyond Sophomore Cohort: 1980-1992, Postsecondary Transcripts (NCES, 1995), he tracked a national sample of high school sophomores from 1980 until 1993. Roughly 65% of his sample of approximately 13,000 attended some form of post-secondary education, and 40% attended a four-year college or university. Additionally, 63% of that sample earned bachelors degrees (Introduction, p. 1). Adelman employed multiple linear regression to identify those factors most influential for college graduation. His eleven-variable, statistically significant model accounted for 43% of the variance in bachelor's degree completion (Executive Summary,
A summary logistic regression model, also statistically significant, provided "both a dramatic underscoring of the principal findings and some enlightening variations" (Executive Summary, p. 1). From his analysis, Adelman proposed that several traditional variables related to college completion be discarded, and he constructed others that proved more meaningful. Among those discarded were persistence, measured in terms of students returning to college from one year to the next, because the quality of their persistence was not considered in the measurement. Likewise, "academic track," (p. 5) a descriptor associated with high school transcripts, was judged too variable in terms of curriculum quality to be included in the model. Conversely, for the purposes of this study, one reconstructed variable stands out: intensity of students' high school curricula.

This is the most elaborate construction in the study. It includes Carnegie units in 6 academic areas, accounts for highest mathematics studied, remedial work in English and math, and advanced placement. The construction results in a criterion-referenced scale with 40 gradations. (Executive Summary, p. 5)

Adelman further redefined the mathematics variable associated with academic intensity.

For mathematics, four variables were created: all high school mathematics credits, remedial mathematics units, net mathematics units (all minus remedial), and HIGHMATH, a variable indicating the highest level of mathematics reached by the student in high school. HIGHMATH proved to be an extremely powerful construct.... (Introduction, p. 2)

The intensity variable was included with others to form a composite variable called the "academic resources." A logistic regression model of bachelor's degree attainment by age 30 generated odds ratios from which a partition of resources emerged. The high school curriculum accounted for 41% of those resources, while college entrance test scores accounted for 30%, and class rank/GPA accounted for 28.7% (Part I, p. 9).
Hence, "No matter how one divides the universe of students, the curriculum measure produces a higher percent earning bachelor's degrees than either of the other measures. The correlation [using linear regression] of curriculum with bachelor's degree attainment is also higher (.54) than test scores (.48) or class rank/GPA (.44)" (Executive Summary, p. 2). It is also important to note that the impact of curriculum intensity is even higher for some minorities (Executive summary, p. 2).

In light of his methods and analysis, Adelman reached three conclusions that are both striking and meaningful to admissions officers and to the research at hand.

When the academic intensity and quality of one's high school curriculum is such a dominant determinant of degree completion, and both test scores and (especially) high school grade point average or class rank are so much weaker contributors of attainment, college admissions formulas that emphasize test scores and (especially) high school grade point average or class rank are likely to result in lower degree completion rates. (Executive Summary, p. 4)

Of all pre-college curricula, the highest level of mathematics one studies in secondary school has the strongest continuing influence on bachelor's degree completion. Finishing a course beyond the level of Algebra 2 (for example, trigonometry or pre-calculus) more than doubles the odds that a student who enters post-secondary education will complete a bachelor's degree. (Executive Summary, p. 2)

The most useful data lie in the details, not the generalities. (Executive Summary, p. 5)

Although the first two of the above conclusions are clear, the third begs further explanation. In remarks further criticizing the "X-Percent' Solution" (Part V, p. 3), Adelman strongly prescribes measures different from those in common use by admissions officers.

And yet it is to high school class rank and GPA that policy makers have turned for cheap and easy 'solutions' in admissions to public institutions that exercise any degree of selectivity... A Test score is a snapshot of
performance on a Saturday morning. Secondary school grades-and the relative standing that they produce in "classes" where the student body may be constantly changing-carry as much reliability as a pair of dice (Elliot and Sterna, 1998). But the intensity and quality of curriculum is a cumulative investment of years of effort by schools, teachers, and students, and provides momentum into higher education and beyond. It obviously pays off. The effects of grades and test diminish with time, but the stuff of learning does not go away. (Part V, p. 3)

A methodological note is in order regarding Adelman's second conclusion about his "HIGHMATH" variable as best predictor of success. That researcher's logistical model produced odds ratios for each of his independent variables as predictors, and HIGHMATH had the highest ratio of 2.59 ($t=14.1, p > .0001$). The ratio applies to each of five levels of mathematics curricula: calculus, pre-calculus, trigonometry, algebra 2, and below algebra 2, in descending order. That is, at each increasing level, passing students were 2.59 times more likely to complete a bachelor's degree (Part I, p. 6).

In conclusion, it is appropriate to note that Adelman's research counted traditional mathematics courses only, since the reform movement had not yet begun. His data on high school performance predated the NCTM Standards by seven years. While curriculum intensity may be the ideal measure associated with college completion, that measure would not apply to the study at hand, since a reform-based curriculum is fundamentally different from one that is traditional. Also, Adelman affirmed the relatively high correlation between college admissions tests and bachelor's degree attainment (.48), which is, nevertheless, a solid statistic, and one which this researcher will use to defend his own use of ACT and SAT scores. Very compelling, however, is Adelman's argument, based on evidence and the disaggregation of data, that speaks more accurately about students' abilities than scores derived from traditional sources.
McCormick (1999) also used data from the High School and Beyond study to generate a multiple linear regression model that included degree attainment as a function of freshman grade point average. His study accounted for 58,000 sophomores attending college between 1982 and 1990 (p. 57). The researcher concluded that "Academic performance during the first year was strongly correlated with degree completion: the higher the student's first year GPA, the more likely that student was to have received a bachelor's degree" (p. 6). Although McCormick did not state the correlation to substantiate his remarks, he did provide data (Table 6, p. 22) that clearly indicates increasing frequencies of graduation as first-year grade-point increases. Hence, this result affirmed the use of the freshman GPA in the research at hand.

Multiple sources reported that ACT and SAT scores were commonly associated with first year success in college, with correlation coefficients in the .50 to .60 range, depending on disaggregation (ACT, Inc., 2000; College Entrance Examination Board, 2000; Noble & Sawyer, 1988). Noble and Sawyer cautioned that the combined ACT score was a better predictor than a single score (p. 15), and that each institution needed to develop its own predictive equation using local data (p. 46). Murtaugh, Burns, and Schuster (1999) used regression analysis to create a survival model of retention for 8,867 Oregon undergraduates from 1991 to 1996. Their model indicated that student retention increased with both grade point average and with SAT scores in freshman and sophomore years (p. 356).

ACT and SAT mathematics scores were important in this study, because they determined student placement in mathematics classes, reflected achievement, and predicted success. While the previously cited works used composite ACT and SAT
scores as dependent variables, other studies relied on the mathematics portions only.

Benefield (1996) examined data from 2,505 students in pre-engineering programs at Auburn for fall enrollments in 1991, and 1993-1995. He found that mathematics correlation with completion of the program to be .38 with a mean ACT mathematics score of 26.45 (p. 5). Additionally, the researcher claimed a stronger correlation of .55 between first quarter GPA and completion of the pre-engineering program (p. 6).

Other predictors of college success exist also. Levin and Wyckoff (1990) examined persistence of 1,043 engineering students into their sophomore year at Pennsylvania State University. His logistic regression model suggested that high school GPA to be among other best predictors of persistence along with mathematics placement scores, gender, chemistry and non-science scores, and reasons for choosing engineering as a major (p. 27). Those researchers found that the best freshman year predictors to be grades in Physics I, Calculus I, and Chemistry I, noting that "the pre-enrollment variables are replaced by performance variables in later models" (p. 33). Thus, the best predictors for persistence evolved with the students, according to these researchers.

Adelman (1998) again used data from the National Center for Educational Statistics High School and Beyond/Sophomore Cohort Longitudinal Study (NCES, 1995) to identify best predictors of persistence and success for students in engineering programs. His analyses included between 12,640 and 14,825 records, and he noted that "no other national data sets have such inclusive data" (p. 17). Adelman concluded that on average, mathematics credits comprised nearly one seventh of the total credits taken by engineering students, and that they take more mathematics than any other subject (p. 9). The researcher stated a correlation of .525 (p < .001) between the highest high school
mathematics course taken and completion of a bachelor's degree in engineering (p. 42). He again emphasized that the highest level of mathematics studied in high school was strongly correlated with degree completion in any field, referring to the strong odds ratio in his previously cited work (Adelman, 1999).

In addition to ACT, SAT, and GPA scores, students' initial choice of majors were pertinent to the work at hand. Dawson-Threat & Huba (1996) referred to the psychosocial and developmental theory of Chickering, suggesting that, students "clarify purpose" when they declare majors (p. 297). That is, they set goals and directions while prioritizing and formulating action plans for their futures. The act of clarifying purpose is integrative and forward-looking, because it includes vocational, avocational, recreational, and other life-style aspirations and issues (Chickering, 1969, p. 108). Of course, Chickering's theory may not apply in the case of frivolous or cavalier choices.

Vetter (1994) argued that major choices are formulated much earlier than high school, with critical junctures at nearly every grade level. She noted that the early experiences in elementary school were important in turning students toward or away from science and mathematics. Furthermore, dropping out of high school mathematics constituted a major barrier to learning science. (p. 34) Citing the National Research Council's (1989) findings, she noted that half of all American students dropped out of mathematics by ninth grade. Half of the remainder dropped out each year thereafter up to the Ph.D. level (p. 25). Vetter also noted that National Assessment of Educational Progress (NAEP) reports indicated that the reasoning ability of American students was lower than expected. The numbers in high school science served as indicators of declining interest, according to Vetter's analysis of the NRC report. Although 98% of
high school students took biology, only 49% took chemistry, and 21% took physics (p.
33).

Strenta (1994) constructed a logistic regression model to investigate initial interest
in and attrition from natural science and engineering programs at four highly selective
institutions. From a large sample of students (n=5,320), he was able to isolate and rank
order high school achievement indicators that predict interest in science courses. The
researcher used β-weights for that ranking. The highest predictor variable was high
school grades in mathematics and science (β=1.44), followed by the number of
mathematics and science course taken by students (β= .48) and SAT mathematics scores
(β=.04). Negative predictors from strongest to weakest included high school non-science
grades (β= -.35), gender (β= -.18), and SAT verbal score (β= -.07). SAT scores and
gender were only slightly influential. All predictors were significant with p<.05 for the
gender factor and p<.001 for the others. Achievement test scores (β= .03) were only
slightly influential with at p< .01. It is critical to note that using β-scores to rank the
importance of variables is not meaningful unless the units for each variable are identical.
The researcher did not provide information about transformations that would allow this
type of comparison.

How Reformed Are Mathematics Professors? LaBerge, Zollman, and Sons
(1997) interviewed thirty mathematics faculty at three universities, five four-year
colleges, and three two-year colleges to profile knowledge, beliefs, and practices
regarding professional standards for mathematics education. That research team
preceded their study by asking deans and department heads to suggest participants.
Hence, randomness was immediately suspect, although the authors noted that faculty were selected from a variety of backgrounds (p. 4). The interviews lasted approximately an hour and consisted of four parts: views about mathematics and mathematics education, beliefs about the NCTM Standards, practices in and outside the classroom, and general awareness of professional standards.

LaBerge, Zollman, and Sons reported a low level of awareness of the NCTM Standards by faculty, although 27 of 30 had heard of them. Awareness of the second and third volumes on teaching and assessment, respectively, was lower than for the first published Curriculum and Evaluation Standards. Fourteen faculty members said they were not aware of the Professional Standards for Teaching, and eleven said they were unaware of the Assessment Standards. Faculty awareness for other professional standards was higher, however. More than half reported awareness of early reform documents targeted at college and university departments such as A Call for Change (Leitzel, 1991) and Reshaping College Mathematics (Steen, 1989).

LaBerge, Zollman, and Sons found that faculty "generally agreed with the underlying assumptions of the NCTM Curriculum and Evaluation Standards" (p. 8). Mean scores from the Standards Belief Inventory ranged from 3.80 to 5.55 on a scale of 1.00 to 6.00. A score of 1.00 indicated strong disagreement with a particular standards-based idea and 6.00 indicated strong agreement. On 13 of 20 standards-based items, more than half of the faculty expressed more agreement than disagreement (p. 22).

Twenty-five percent or more of the faculty expressed disagreement on seven standards-based tenets in three clusters. In the "learning" cluster faculty disagreed that all students could learn mathematics and that all students bring with them considerable
experience and at least partial understanding of the subject. In the teaching cluster, faculty believed that a lack of computational mastery should prevent students from proceeding to a subsequent class and that calculators should not be available to all students at all times. Additionally, about half of the faculty believed that early use of calculators would inhibit learning of basic mechanical skills. Finally, 11 of 30 faculty members thought that mathematics "should be thought of as a collection of concepts, skills, and algorithms" (p. 22).

In a third piece of research, LaBerge, Zollman, and Sons used self-reporting to measure the pedagogical practices of their interviewees. The instrument they employed was based on the Classroom Practice Inventory (CPI), adapted from the work of Weiss et al. (See the CPI in APPENDIX B.) A total of twelve student activities are listed on the CPI, nine aligning with recommendations of various sets of professional standards, and three, representing more traditional learning activities. The more traditional activities include working exercises or problems from the text, working problems from teacher-prepared materials, and taking notes in class (p. 10).

Most (28 of 30) faculty members reported having their students take notes during class "most of the time" or "always." At least a quarter of the faculty indicated that their students participated in the following standards-based activities either "most of the time" or "all of the time": using calculators, learning through real-life applications, presenting or discussing solutions to mathematics problems, working exercises or problems from the text, and making conjectures and exploring problem solving methods (p. 10). As the authors note, a third of the faculty used all of the activities sometimes except one: writing about mathematics; twenty-three faculty members "seldom" or "never" involved students
in this activity. Nineteen faculty seldom or never involved students in cooperative
learning and nineteen seldom or never used computers in the classroom. Eighteen faculty
seldom or never used physical materials or models in class, and the same number
indicated that their students seldom or never work on projects or open-ended problems (p.
10).

The authors note that some faculty classified several of the activities as upper
division and lower division. More likely to be classified as upper division were the
following: using physical materials or models, learning through real-life applications,
working on projects or open-ended investigations, and presenting or discussing solutions.
Conversely, working teacher-prepared worksheets and problems was more likely to be
considered a lower-division activity. In the survey, faculty also had the opportunity to
comment on which activities they prescribed for work outside of class. Working
problems from the text was the most common, and 28 of 29 respondents prescribed this
activity "always" or "most of the time." The use of calculators outside of class was also
popularly prescribed by 18 of 28 faculty. Most faculty members indicated that they
"seldom" or "never" expected students to use physical models (19 of 32), write about
mathematics (19 of 26), or present or discuss solutions to mathematics problems outside
of class (14 of 24).

In general the interviewees expressed a desire for students to participate more
often in activities like those espoused in the standards, but they (21 of 30) also indicated
that time constraints prevented them from doing more standards-based activities during
class. Some indicated that these might be better left for after class. The authors also state
that other barriers to implementing standards-based practices, according to faculty
comments, include lack of resources (12 of 30), class size (11 of 30), curriculum content (9 of 30), students' negative beliefs and expectations related to mathematics (7 of 30), a lack of information about activities (6 of 30), and testing or assessment concerns (6 of 30) (p. 20). Other barriers listed include discomfort with new methods, low ability level, poor study skills, and lack of interest and motivation on the part of students, scheduling problems, and the fact that teaching was not valued enough by the institution to warrant extra effort. No numbers were provided for these categories of responses.

The three researchers came to several conclusions through their work with college and university faculty. Faculty are more aware of reform publications that directly inform their educational levels, such as those of the MAA, than they are of those that inform secondary education. To the extent that they are aware of the NCTM standards, they have a favorable impression of them. (Two respondents indicated otherwise.) Moreover, the beliefs of many faculty members align with the assumptions and prescriptions of the NCTM Standards. Classroom practices, while still traditional, are changing, but barriers to change exist, the perception of too much material to cover in too little time being the greatest barrier. LaBerge, Zollman, and Sons concluded with an important observation about how change begins.

The first stage in improving mathematics instruction at any level is awareness of current teaching practices, followed by an acknowledgment that there may be better alternative methods. This sense that there is a reason to change, whether it is a level of dissatisfaction with current practice or a belief that these alternatives can provide better learning for students, must outweigh the risk involved with change. Following initial experimentation with alternative methods, continued change requires further reflection, discussion, mentoring, and a meaningful level of success. (p. 16)
Finally, among the questions that these researchers raised, was the following. "Will students whose learning experiences reflect the Standards have different beliefs about mathematics?" From students' comments in both the Core-Plus and SIMMS studies, it is clear that they already do.

More data about the practices of college instructors in developmental programs came from the dissertation of Kull (1996), who sought a model for developmental programs. To gauge the awareness and incorporation of the NCTM Standards, Kull examined the faculty in 23 self-contained developmental programs. Those programs comprised 65% of the originally chosen sample that represented a population of 464 programs. Forty-eight complete or partially complete surveys were returned from these programs, all based at four-year colleges and universities in the U.S. A committee of four mathematics educators screened Kull's questionnaire for content validity and the test for reliability (test-retest method) yielded a coefficient of .83. Reliability coefficients for each question ranged between .54 to .96. Kull speculated that higher coefficients reflected those clearly understood practices, such as journal writing and writing computer programs, while lower coefficients indicated less specific methods, such as the use of problem solving or modeling.

Kull's research design was descriptive in nature. 62% of the respondents reported having read the NCTM Curriculum and Evaluation Standards for School Mathematics, while 46% stated that they had read the Professional Standards for Teaching Mathematics. A typographical error in the original classroom survey forced the researcher to mail corrected forms to respondents, resulting in the return of two types of completed instruments (p. 127). Despite that glitch, conclusions about classroom practice
were clear from Kull's tabulations. More than 50% of the instructors replied that they never used manipulatives, that they never required students to keep journals, collect and process data, or use computers as part of their problem-solving processes. Also, most instructors never required student-generated computer programs or group presentations, nor did they use student-generated problems for evaluations. Conversely, more than 50% of the instructors required students to use calculators, taught problem-solving skills, and used instructor-guided whole-class discussions. Instructors seemed to find a "balance" (p. 144) in three teaching methods: student-led class discussions, individual projects and research, and the use of calculators to explore relations and functions. Kull concludes that "some instructional and evaluative activities [were] going on in developmental mathematics classrooms that are envisioned by the Standards," but the nature of his descriptive research design made it impossible to attribute these practices to the NCTM Standards (p. 145).

Enrollment in remedial mathematics programs deserves more consideration in light of the work by Kull (1996), considering that one component of the study at hand is concerned in part with developmental courses at Montana universities. Lewis (1996) surveyed 3,060 two and four-year colleges to gather information about remedial programs, finding developmental courses to be pervasive in American colleges and universities. On the average the institutions profiled offered 2.5 remedial mathematics courses, higher than both the 2.1 course average for remedial reading and the 2.0 course average for remedial writing (p. 6). The mean pass rate for remedial mathematics courses was 74% among those institutions surveyed (p. 6), and students who took remedial courses had lower second year return rates than those who did not take them (p. 23). In
addition, 66% of surveyed institutions used their remedial mathematics courses as
gateways, requiring them to be completed as prerequisites. While enrollment rates for
remedial reading and writing courses did not significantly differ between 1989 and 1995,
enrollment rates for remedial mathematics did, from 21% to 24% (p. 44). (The author did
not provide significance data.) By virtue of their numbers and qualities, then, remedial
courses deserved to be considered separately in any analysis, such as that provided in this
study.

Summary of Methodologies

Introduction. The methodologies important to this study included those
associated with research in three areas: the meaning of college-related performance
measures, research models for college-related measures, and standards-based practices in
the college mathematics classroom. College-related measures of interest included ACT
and SAT scores, college freshman grades, and initial choice of major. The research
models compared were those of Bachelis (1998) and Milgram (1999), the Core-Plus
evaluation (Core-Plus Mathematics Project, 2000b), and the SIMMS freshman
evaluations (Allinger, Lott, & Lundin, 1998). From those studies sprang the research
model for this study. Finally, methodologies from the work of LaBerge, Zollman, and
Sons and from Kull provided information about standards-based practice in college
classrooms. In particular from the former trio of researchers' came the Classroom
Practices Inventory used in this study. Methodologies associated with college-related
measures, research models, and practice in the classroom have been carefully considered
in the previous sections, and are summarized and compared in what follows.
College-related Measures. Several large studies, cited previously, related SAT and ACT scores to freshman grades and to degree completion via linear and logistic regression models. Adelman (1998, 1999) and McCormick (1999), using these modeling methods on nationally representative data, isolated important dependent variables for predicting achievement in college, as did Murtaugh, Burns, and Schuster (1999) on a smaller scale at an Oregon university. Likewise, Noble and Sawyer (1988) produced results independent of ACT, Inc. and the College Entrance Examination Board (SAT) that show high correlation between these tests and freshman grades.

When only the mathematics portions of the ACT or SAT tests were considered, the linear regression model of Benefield (1996) demonstrated that mathematics ACT scores were highly correlated to completion of a pre-engineering program, as were first quarter GPAs. An older regression model of Levin and Wyckoff (1990) also demonstrated the predictive quality of ACT scores. Dorans, Felicia, Pommerich, and Houston (1997) established strong correlation between SAT and ACT scores, generating concordance tables for the two tests. Lewis (1996) established the pervasive nature of remedial mathematics in American colleges and universities, from which this researcher infers the need to treat results of developmental courses separately from others. In conclusion, as a result of large regression models, the prediction of freshman college performance by SAT and ACT scores has been well-documented, as has the prediction of degree completion from first year (and subsequent years’) performance. The cited research substantiates the use of ACT and SAT mathematics scores and freshman mathematics grades in the study at hand.
Initial choices of majors are an important measure for this study. Dawson-Threat and Huba (1996) grounded students' choices of majors in Chickering's psychosocial theory, while Vetter (1994) reasoned that student attrition from mathematics and science begins well before college. Stranta's (1994) logistic regression model confirmed that researcher's hypothesis: Interest in and attrition from mathematics and science courses was predicted by grades in those subjects and by the number of courses taken in high school. Regarding the study at hand, counting SMET majors, then, was a strategy supported by psychosocial theory and by empirical evidence, linking interest in science and mathematics, high school performance, and choice of majors.

To gauge the intensity of reform of college mathematics professors, the work of LaBerge, Zollman, and Sons (1997) was indispensable, since it provided the instrument for this study. Their work, largely descriptive, combined with that descriptive research of Kull (1996) and the sophisticated causal model of Hawkins (1998) to paint a portrait of mathematics reform in college classrooms. Notably, the content measured by a portion of the survey of LaBerge, Zollman, and Sons was very consistent with that found in Kull's instrument. The work of Hawkins affirms that standards-based practice positively affects college students' self-concepts as well as their interest and achievement in mathematics.

The Core-Plus Study. Bachelis (1998) and Milgram (1999) used a questionnaire in their tandem study of the Core-Plus curriculum, and their problem was well-stated. Variables, including ACT scores, SAT scores, and course grades, were implied rather than being explicitly stated. Neither author provided a formal review of the literature, although both authors issued a brief history of the Core-Plus program and a list of
references for citations. The authors did not formally state hypotheses, but these came out in the context of the article. Bachelis did not write that he tested his survey instrument for either validity or reliability, and, since most of the analysis was done by Milgram, it is not clear how much was planned or merely included post facto. Bachelis did not attempt to avoid leading questions in the survey, and neither the instrument nor the cover letter to potential respondents seemed to contain biasing remarks.

The greatest flaws in the research of Bachelis and Milgram were in their methods of sampling and data gathering. While they mailed surveys to all of the students from the Core-Plus and non-Core-Plus high schools, their return rates, 50% and 30% respectively, were well under those prescribed for either population. (Gay, 1996, p. 125) The default sampling process was self-selection, and the self-reporting of all scores in the study weakened credibility of the study. While this study qualified as a glorified pilot, the Bachelis and Milgram work cannot stand up to professional scrutiny, and any conclusions were suspect. Considered as a pilot, however, their work emphatically pointed to variables that merited consideration in more rigorous studies. ACT and SAT scores and freshman grades, for example, were considered in the work at hand. Notably, the work of Bachelis and Milgram captured the tensions between reform-based and traditional tenets and values, as it tacitly pointed to a need for tighter research methodology.

The SIMMS Study. Allinger, Lott, and Lundin (1998) employed a two-stage method in their study entitled "Attitudes and Performance of College Freshmen Who Used the SIMMS Integrated Mathematics Curriculum." Stage one involved interviewing 33 of 66 former SIMMS students in their second semester of college. Those interviewees
were chosen using a modified stratified sampling process that spanned schools represented by the SIMMS project. Stage two involved a Fall 1997 comparison of former SIMMS students to all traditional freshmen at six universities and colleges in Montana. That comparison included mathematics placement scores and pass rates in mathematics courses (Allinger, Lott, and Lundin, 1998, p. 19).

That study was not a refereed work but was published in the SIMMS Monograph 5: The Classroom (1998). As such, it did not have a literature review or a list of references. A formal problem statement was included in the introduction (p. 16), however, although no hypothesis is stated formally. Variables were not formally defined but were implied in the description of the study. Sizes and major characteristics of populations were well-described, and all non-interview data came from the official records at six higher education institutions. There was no mention of testing the interview instrument for either reliability or validity, and no pilot study was conducted before the interviews. Although some interview questions could be considered leading, these were balanced by questions potentially leading in the opposite direction. The interview protocol did not control for prompting or probing, and individual responses were typed or written. Interviews lasted from 15 to 35 minutes. The researchers noted that "considerable effort was made to reflect the actual quantity and quality of student comments without bias for or against the [SIMMS] Project" (p. 19). These interviews were quite structured, so that codified responses generally fit the protocol.

Although not a refereed research paper, the SIMMS study was careful enough to be credible. Despite the researchers' involvement with the SIMMS project, they included a number of student responses critical of the SIMMS curriculum, highlighting in
particular the issue of algebra as a potential barrier for SIMMS students. Moreover, they included statistics to indicate course types where those students did not perform as well, proportionately, compared to the general traditional freshman population. The work of Allinger, Lott, and Lundin should be considered a strong pilot leading to the work at hand. This study will include inferences and tests for significance. When the work of Allinger, Lott, and Lundin was compared with the Core-Plus study, important measures became well-identified. In particular, the disaggregation of freshman course grades into developmental, courses before calculus, and calculus became important, as did freshman grades and college entrance examination scores.

The Faculty Studies. The study of LaBerge, Zollman, and Sons (1997), highlighting the standards-based awareness, beliefs, and practices of college and university faculty, was accepted for publication in the journal, Focus on Learning Problems in Mathematics. As such, it is one of two works—the other study comes from Kull (1996)—that numerous literature searches produced, which quantified the views and practices concerning professional standards of college mathematics faculty. The study of LaBerge, Zollman, and Sons contained a comprehensive review of the literature, although that review needed updating, as did, possibly, the theoretical framework. The research topic was well-stated in the title, and implied elsewhere. Participant selection is well-described, but the selection process suffers, because department heads or deans were asked to suggest names for participants in the study.

The interview format had four parts. Part one was designed to draw out the views of faculty members. Part two measured the beliefs of faculty using the Standards Beliefs
Instrument (SBI). The SBI was tested for construct validity by an expert panel (chi-squared test) and checked for reliability with internal consistency using Spearman-Brown tests, and the authors provided statistics for both. Part three of the interview was conducted with the twelve-question Classroom Practice Inventory (CPI) adapted from Weiss, Upton, and Nelson (1992). The adaptation itself was never tested for reliability, although all questions later appeared in a nationwide study by Weiss, Matti, and Smith (1994), that was extensively validated. Also, many of the questions were remarkably similar to those found on Kull's (1996) instrument, which was both validated and tested for reliability. The brief fourth part of the interview asked faculty about their awareness of the standards, and no data on validity was given.

A complete data output was included in the appendix of that paper, and the conclusions drawn by the authors seemed to match the questions asked and the data generated. The authors did not describe the interviewers or the method of recording data, but their discussions were complete enough to indicate unbiased responses of the interviewees.

It is unfortunate that its original creators did not report testing the Classroom Practices Inventory (CPI) for validity. The CPI has weathered some use, however, and gained credibility despite this. Its format is simple, and the questions are direct, perhaps leading previous researchers to conclude that the instrument is both valid and reliable, despite a lack of statistical conformation.

Kull's (1996) work on the alignment of self-contained developmental mathematics programs with the NCTM standards produced a classroom practices inventory that was formally validated and checked for reliability. In addition, the instrument included
questions on assessment and evaluation, and so was more complete and dependable than the instrument used by LaBerge, Zollman, and Sons. Importantly, the content of Kull's instrument was very much like that of the former trio of researchers, further affirming that either instrument would measure classroom practices in a valid and reliable manner. Also, Kull's findings in the developmental setting mirrored those of LaBerge, Zollman, and Sons: College instructors were moderately aware of the NCTM Standards, used some standards-based classroom practices, but did not employ others. Those practices employed by instructors in both studies had similar relative frequencies, indicating selective standards-based practice.

Conclusions

Summary of the Review of the Literature

How university freshman, as veterans the SIMMS Project's high school mathematics curriculum, compared with those who used a non-SIMMS curriculum needed to be considered within the rich, if not colorful, context of the mathematics reform. The reform movement is in flux, and tensions between reformist and traditional camps do not remain inside the professional arena. Various professional standards, as well as guidelines from a major grantor, NSF, evolved with the movement. Ironically the reform effort may have suffered from mutual adaptation, a recommended planned change process, because evaluation of college-related measures did not remain in the list evaluation requirements of NSF.
The review of literature outlined large studies linking college entrance examination scores to first year college performance and first year performance to graduation. Other studies showed strong relationships among scores on ACT and SAT examinations, first year grades, and completion of specialized, math-dependent programs. While other predictors of college success, such as intensity of high school mathematics programs, may be stronger, they are not employed here, because research supporting this data involved only traditional mathematics curricula.

Literature on the effects of reform-based curricula on college students, although sparse and weak, is methodologically suggestive. The Core-Plus and SIMMS studies indicated that students coming from reform-based curricula faced algebraic barriers when placing into some college freshman mathematics classes. College entrance test scores reflected an algebraic disadvantage for those reform veterans, since entrance tests tend to be more traditional algebraic content. Self-reported data and self-selective sampling marred the Bachelis and Milgram study, which employed a test for significance in the placement scores of students from both camps. Other contradictions ensued when information from Core-Plus' own evaluation was factored into the analysis. Although there was evidence that SIMMS students and Core-Plus students underwent forced paradigm shifts when exposed to college mathematics courses, there was also evidence that general identity-crisis, resulting from novel college experiences, occurred among those students from more traditional backgrounds. This dynamic is, in fact, necessary for growth, according to psychosocial theories (Evans, Forney, & Guido-DiBrito, 1998, p. 36; Pascarella & Terenzini, 1991, p. 19). Indicated, then, was the need for more comparative research and, especially, for more rigorous methodology in comparisons of
reform-based mathematics curricula to those that are more traditional. That was a goal of the study at hand: To rigorously compare ACT and SAT scores and freshman mathematics grades of students from SIMMS and non-SIMMS backgrounds. Indicated also, was the need to find out if students from reform-based curricula initially choose math and science majors differently from their traditionally educated cohorts. That too was a goal of the study at hand.

The two analyses of college mathematics faculty (LaBerge, Zollman, & Sons, 1997; Kull, 1996) provided another context in which to view incoming freshmen, and this context was more traditional than reformed. Faculty tended to favor practices such as taking notes, mastering computation, and doing textbook problems, although those faculty seem to want to change. Comments from the Core-Plus and SIMMS studies suggested that students feel pressure to switch paradigms upon entering college classrooms or suffer the consequences. As previously noted, those students in transition from a reform-based high school program to traditional college mathematics classes experienced "crises" in the sense of Erickson (Pascarella and Terenzini, 1991, p. 19). Resolving those crises, however, might depend on how students were treated in college classrooms. Hence, a third goal of this study was to determine if reform-based classroom practice differed among three Course Types: Developmental, mathematics Before Calculus, and Calculus.

Weaknesses and Strengths in the Literature

Literature on the effects of reform-based mathematics curricula on college students was meager. The curricula are relatively novel, and there has been limited
pressure from NSF to conduct college performance evaluations. Also, systems of articulation between secondary and post-secondary schools did not seem to encourage this type of assessment. Despite methodological weaknesses, the Core-Plus study pointed to tensions between reform-based and traditional practices, as did the SIMMS study. Together, these studies pointed out possible barriers for students caught between educational paradigms as well as to methodologies needed to better evaluate and compare programs. However, it was not established whether the veterans of reform-based projects differed significantly from veterans of more traditional mathematics programs. That was the business of the current study, which focused on comparing one such group of veterans, former SIMMS students, to their more traditionally educated cohorts.

Ample and sound evidence existed, linking ACT and SAT scores, freshman grades, and more general college performance, such as that from Adelman (1998; 1999) and McCormick (1992), and Murtaugh (1999). The findings repeatedly affirmed those linkages, and, more specifically, they affirmed the correlation of mathematics scores on entrance examinations to college performance (Benefield, 1996; Levin & Wycoff, 1990). In those cases, the studies were large, and their regression methods, sound.

The research on choice of college majors was less ample, but it was coherent. Dawson-Threat and Huba (1996) grounded major choice in psycho-social theory and Vetter (1994) noted early influences on choices of majors. Strenta (1994), using a logistic regression model, affirmed the suspected correlation between interest in mathematics and science at the college level and high school performance. Together, those studies suggested that major choice was linked to students' high school experiences.
The literature search uncovered only two studies of college faculty views and
practices as they aligned with the standards. The studies of LaBerge, Zollman, and Sons
(1997) and Kull (1996) constituted snapshots of faculty awareness, beliefs, and practices
that are now dated. Nevertheless, they suggested that mathematics faculty and instructors
were caught in a shift of pedagogical beliefs and practices associated with mathematics
reform. The Classroom Practice Inventory, which was used in the Instructor Survey of
the present study, came directly from the work of LaBerge, Zollman, and Sons.
However, while its validity was ascertained, there was evidence of the CPI being tested
for reliability, although the content of the instrument was very similar to that used by
Kull (1996), which proved reliable.

Gaps and Saturation Points in the Literature

Repeated evidence in the literature supported the use of ACT and SAT scores to
predict performance in college and the validity of doing so. Likewise, freshman grades
were linked to later performance and graduation. From the literature review, it was
evident that both public and scholarly interest existed in traditional evaluations of
college-related measures: ACT and SAT scores and freshman grade point averages are
measures that turned up often.

On the other hand, only the work of Vetter (1994) and Strenta (1994) alluded to
the effects of high school curriculum on major choice; no other studies were found on
this subject. No research-based model was uncovered that tests reform-based effects on
choice of college majors, although original NSF Driver Five called for this type of
evaluation.
Much more could have been written about the nature of the SAT and ACT tests, and about the constructs they measure. It was assumed with some justification that both ACT and SAT tests were of a more traditional nature, because they do not and cannot agree with authentic assessment concepts embedded in the various sets of professional standards.

The calculus reform movement saw its own evolution during the last ten years, but the review of literature did not included any studies in this area. Research in calculus reform was beginning to highlight the benefits of this approach, however. In the meta-research article, "Effects of Calculus Réform: Local and National," Hurley, Koehn, and Gantner (1999) address the many positive outcomes enjoyed by a host of such projects. Some of these projects have managed to bypass the algebra barrier by demonstrating that students taking a reformed calculus course outperform those in a more traditional course, despite traditional testing (p. 803). There are obvious parallels between the secondary and post-secondary reform projects, but these are beyond the scope of this study.

Avenues of Further Inquiry

The literature review brought to the surface qualitative and descriptive data that begged for more precise methodology to answer the original research question. Do former SIMMS students differ from their more traditionally educated peers in college-related measures such as SAT and ACT scores, freshman mathematics grades, and initial choices of majors? The SIMMS study suggested that data on course grades be disaggregated by Course Type: Developmental, courses Before Calculus, and Calculus. The Core-Plus, SIMMS, and Faculty studies indicated tensions between educational
paradigms that further merited this disaggregation to discover differences in student performance in those levels.

Much work beyond this study is needed to determine the extent to which both secondary and post-secondary mathematics departments comply with professional standards, especially where program evaluations are concerned. The literature on change, on assessment and evaluation, and the standards themselves outline the necessity for effective program evaluations to maintain program quality. A first effort in that direction would involve surveying high school and college mathematics departments, measuring needs, beliefs, and practices at both levels.

Assessments that measure the abilities of students from the standpoint of professional standards must occur if, as claim Romberg and Wilson (1995), mathematics is "a dynamic set of interconnected, humanly constructed ideas" (p. 4). While the study at hand focuses on the more traditional measures, those sought by Romberg and Wilson do not. The assessments they propose focus on those standards-based elements that continue to make mathematics interesting, dynamic, vital, and popular among mathematicians themselves. There can be no double standard; mathematics education at all levels deserves to bask in the qualities that mathematicians experience in their beloved subject. Much work awaits researchers in this area.

Chapter Summary

This review presented, from a change theoretic perspective, an evolution of various sets of standards and drivers for mathematics reform movement. The review provided a basis from which to do research toward answering the main research question:
How do former SIMMS students compare with their more traditionally educated peers in college related measures? The Core-Plus and SIMMS studies supplied information about the effects of reform-based curricula on college students. While the Core-Plus study brought forth important criteria for comparison, such as ACT and SAT scores, course grades, and course placement rates, the SIMMS study more rigorously linked those same criteria to the work at hand. Within the theme of change, the literature brought out tensions between the reform-based paradigm and one that is more traditional, and evinced the need to focus on students and their struggles to reconcile two seemingly competitive systems. While the Core-Plus study lacked rigor and the SIMMS study lacked tests for significance, these two works pointed the way down a more structured research path.

Various studies on college-related performance strongly linked ACT and SAT scores to freshman grades and further performance. Major choices were grounded in psychosocial theory and related to interest and earlier performance. The faculty studies provided an instrument with which to gauge departmental alignment to standards-based practice, so that any conclusions about freshman performance could be measured against what actually happens in freshman mathematics classes.

There is a dearth of literature on the effects of mathematics reform on college students, but this literature is now beginning to appear. More could be said about what ACT and SAT scores measure and the link between calculus reform and reform in secondary mathematics. More needs to be said about program evaluation and its effect on program quality and about assessments that measure what is really important about mathematics. This study, though, considered more traditional college-related measures, ACT and SAT scores and freshman mathematics grades, because of their documented
predictive value. It considered initial choices of majors, because they are an indicator of past interest in mathematics. Finally, the Instructor Survey and Supervisor Interviews provide a backdrop for interpretation of freshman performance and a profile of reform in the college classroom. Methodologies involved in these research components are considered next.
RESEARCH METHODOLOGY

Chapter Introduction

This chapter describes the research methodologies used to determine if former SIMMS students, as entering freshmen at two state universities in Montana, differed from their more traditionally educated peers in the college-related performance measures of ACT and SAT scores, freshman mathematics grades, and initial choice of majors. For the purpose this study, the researcher referred to University E and University W as the two institutions of interest. There were three research components described herein: the Freshman Data component, the Instructor Survey component, and the Supervisor Interview component.

This study is descriptive in the sense of Gay (1996, p. 249), who defined descriptive research as "collecting data in order to test hypotheses or to answer questions concerning the current status of the subject of the study" (p. 249). For the Freshman Data component, former SIMMS and non-SIMMS students were compared on ACT and SAT mathematics scores, SAT Verbal scores, ACT English, Science Reasoning, and Reading scores, first-year college grades, and initial choices of majors. The two groups' freshman grades were disaggregated according to three Course Types: Developmental, mathematics Before Calculus, and Calculus. For the Instructor Survey, the Classroom Practices Inventory (APPENDIX A) was used to gauge the classroom practices of college
mathematics instructors teaching courses of the same thee types. The latter research component was conceptually, but not statistically, linked to the analysis of freshman grades according to Course Type.

A third research piece, the Supervisor Interviews, consisted of interviews with eight freshman mathematics course supervisors at each of the universities of this study. Those interviews, guided by an Interview Protocol (APPENDIX A), added depth, dimension, and affirmation to results gleaned from the first two pieces. In those interviews, course supervisors were asked about the goals, objectives, standards, and outcomes they had for the students in their courses. Their responses were categorized according to emergent themes and the results were merged into the analyses of the Freshman Data and Instructor Surveys.

The remainder of this chapter was devoted to defining the participants in this study, the instruments and materials used, and the research design. Additionally, the assumptions and limitations in the INTRODUCTION TO THE STUDY were expanded and followed by a Timeframe and Chapter Summary.

Participants

Population and Subjects

Freshman Data. In this study both institutions, University E and University W, were treated separately, as housing two separate populations, the population of former SIMMS students at that institution and the population of non-SIMMS students there. Each of those two populations consisted of freshmen who were Montana high school
graduates in 1998, but the two populations were distinguished from one another on the basis of subjects' experience with the SIMMS curriculum. Those freshman students who were known to be from SIMMS participating schools and those who are from dual track schools, but had either "SIMMS" or "Integrated Mathematics" on their high school transcripts, were categorized as "SIMMS students" for this study, if they had three or more years of SIMMS mathematics. Conversely, those who were not from known SIMMS schools or were from dual track schools, but who did not have either SIMMS or Integrated Mathematics on their high school transcripts, were categorized as non-SIMMS students. The non-SIMMS group may have contained those students from dual track schools who had fewer than thee years of SIMMS mathematics, as indicated by their high school transcripts. It was presumed that all entering freshmen in this study had passed at least three years of high school mathematics in accordance with the requirements at both state universities in Montana.

Instructor Survey. The Instructor Survey involved gathering data about reform classroom practices from instructors of freshman mathematics courses at both state universities in Montana. Instructors' responses to the survey items were grouped by Course Type: Developmental, Before Calculus, and Calculus. (See APPENDIX C.) All instructors of those Course Types were asked to complete a survey form, and that form linked their responses to a particular Course Type. Hence, there were six potential populations of instructors, three at each institution, corresponding to the three categories of freshman mathematics courses. For self-paced courses, interviews with course
coordinators were used to gather information about practices, since the survey would not be appropriate.

**Supervisor Interviews.** Eight freshman mathematics course supervisors from University E and eight from University W were interviewed about the objectives, goals, standards, and outcomes they had for their students. One supervisor, who was severely ill, suggested that the researcher interview another instructor instead, and his suggestion was followed. The interviewees represented every freshman course taught during the Fall 2000 period of this study.

**Method of Selection of Subjects**

**Freshman Data.** Data for the college-related measures came from SIMMS and non-SIMMS populations of freshman students who had entered both state universities after graduating from Montana high schools in 1998. The data collection strategy consisted of securing data from the registrars at University E and University W and spot checking it against a sample of records available through the SIMMS office. Note that the data solicited from both registrars comprised two complete populations of 1998 Montana high school graduates rather than representative samples.

**Instructor Survey.** All mathematics instructors at both Montana universities teaching courses designated as Developmental, Before Calculus, or Calculus, as defined in this study, were surveyed using the Standards-based College Practice Inventory. If an instructor taught more than one type of freshman course, that instructor was asked to
complete a survey for each Course Type. For self-paced courses, data were collected in the Supervisor Interviews.

**Supervisor Interviews.** Eight course supervisors from each of the two state universities were interviewed. With this selection, an interviewee represented every freshman course at either institution.

**Size, Demographics, and Variables**

**Freshman Data.** 1,407 freshmen from Montana high schools entered one Montana university in Fall 1998 (Quick Facts from 1998-99, 1999), and 1379 entered the other at that time, according to the Institutional Research Staff (C. Burleson, personal communication, February 4, 2000). Not all of these qualified for this study, however, due to late transfers into the state, home schooling, and non-1998 graduation date. The SIMMS staff estimated that 75 to 100 of those freshmen at either institution were former SIMMS students (L. Felix, personal communication, September 30, 1999).

Former SIMMS students came from two types of settings, single track and dual track curriculum implementations. "Single track implementation" means that the SIMMS curriculum was the only curriculum available to a group of students in a particular school, while dual track denotes that another curriculum, besides SIMMS, was available to students. However, despite the availability of another option, SIMMS students must have completed three Carnegie units of SIMMS to be included in that category. Those that did not complete at least three years of SIMMS defaulted for analysis to the non-SIMMS category.
The independent variables for the Freshman Data component were Curriculum Type consisting of two levels, non-SIMMS or SIMMS, and Course Type, consisting of three levels, Developmental, Before Calculus, and Calculus. There were eight dependent variables associated with the performance aspect of the Freshman Data component: SAT Mathematics score, SAT Verbal score, ACT Mathematics score, ACT English score, ACT Science Reasoning score, ACT Reading Score, Fall 1998 mathematics grade, and Spring 1998 mathematics grade. Freshman mathematics course grades were disaggregated according to three Course Types: Developmental, mathematics Before Calculus, and Calculus. Finally, three categories, SMET, Non-SMET, or Undeclared, determine a dependent variable, Major Choice, the measurement of which takes on one of the three levels. More specifically, those three values include science, mathematics, engineering, or technology (SMET) as a major, a major that is none of the previous (non_SMET), or a lack of declaration of a major (Undeclared). Note that that the latter variable, Major Choice, connotes a nominal measurement scale, while the SAT and ACT scores were based on an interval scale, as are freshman course grades.

Instructor Survey. Table I below contains admissible courses, at both Montana universities, that comprise the three Course Types: Developmental, Before Calculus, and Calculus. Table I also contains the number of sections of each course offered in Fall 2000 along with the numbers of instructors for each course. Note that if an instructor taught multiple courses of different types, that instructor was asked to complete a survey for each type of course he or she teaches.
Table 1. Fall 2000 Freshman Mathematics Courses Offered by Institution and Category.

<table>
<thead>
<tr>
<th>Category</th>
<th>Course</th>
<th>Number of sections</th>
<th>Number of instructors</th>
</tr>
</thead>
<tbody>
<tr>
<td>University W</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developmental</td>
<td>U 002 -Prealgebra</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>U 005 -Beginning Algebra</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>U 100 -Intermediate Algebra</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Mathematics</td>
<td>U 107 -Contemp. Mathematics</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Before</td>
<td>U 117 -Prob. and Linear Math</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Calculus</td>
<td>U 130 -Math for Elem. Teachers</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>U 131 -Math for Elem. Teachers</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>U 241 -Elementary Statistics</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Calculus</td>
<td>U 121 -Precalculus</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>U 150 -Applied Calculus</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>U 152 -Calculus I</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>U 153 -Calculus II</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td><strong>Subtotals</strong></td>
<td><strong>74</strong></td>
<td><strong>74</strong></td>
</tr>
<tr>
<td>University E</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developmental</td>
<td>MATH 085 - Prealgebra (COT)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>MATH 103 - Introductory Algebra</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>MATH 105 - Alge. for Col. Students</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Mathematics</td>
<td>MATH 110 - Trigonometry</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Before</td>
<td>MATH 130 - Math-Elem Teachers I</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Calculus</td>
<td>MATH 131 - Math-Elem Teachers II</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>MATH 150 - Liberal Arts Math</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>MATH 151 - Language of Math</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>MATH 216 - Statistics I</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>MATH 217 - Statistics II</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Calculus</td>
<td>MATH 160 - Precalculus</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>MATH 170 - Survey of Calculus</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>MATH 175 - Calculus Technology I</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>MATH 176 - Calculus Technology II</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>MATH 181 - Calculus &amp; Anl Geom I</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>MATH 182 - Calculus &amp; Anl Geom II</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td><strong>Subtotals</strong></td>
<td><strong>70</strong></td>
<td><strong>70</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Totals</strong></td>
<td><strong>144</strong></td>
<td><strong>144</strong></td>
</tr>
</tbody>
</table>

The independent variables for the college Instructor Survey were Course Type, having three levels: Developmental, Before Calculus, and Calculus, and Institution with
two levels: University E and University W. There were twelve dependent variables, consistent with questions on the Classroom Practices Inventory (APPENDIX A). Of those variables, three represented traditional measures of classroom practice, and nine, measures based on recommendations common to secondary and post-secondary professional mathematics standards. These Variables were listed in Table 2 below.

Table 2. Dependent Variables and Corresponding Standards for the College Instructors Survey.

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Traditional:</strong> Measures students' frequency of</td>
<td></td>
</tr>
<tr>
<td>1. working problems from the book</td>
<td>Worthwhile Tasks</td>
</tr>
<tr>
<td>2. working problems from worksheets</td>
<td>Connections</td>
</tr>
<tr>
<td>3. taking notes in class</td>
<td></td>
</tr>
<tr>
<td><strong>Standards-based:</strong> Measures students frequency of</td>
<td></td>
</tr>
<tr>
<td>1. using physical models or materials</td>
<td>Worthwhile Tasks</td>
</tr>
<tr>
<td>2. learning from real-life applications of concepts or procedures</td>
<td>Connections</td>
</tr>
<tr>
<td>3. working in groups</td>
<td>Communication</td>
</tr>
<tr>
<td>4. conjecturing and exploring problem-solving methods</td>
<td>Discourse</td>
</tr>
<tr>
<td>5. using calculators</td>
<td>Problem Solving</td>
</tr>
<tr>
<td>6. working on mathematics projects or open-ended investigations</td>
<td>Reasoning</td>
</tr>
<tr>
<td>7. writing about mathematics</td>
<td>Technology</td>
</tr>
<tr>
<td>8. using computers</td>
<td>Worthwhile Tasks</td>
</tr>
<tr>
<td>9. presenting or discussing solutions to mathematics problems</td>
<td>Problem Solving</td>
</tr>
<tr>
<td></td>
<td>Communication</td>
</tr>
</tbody>
</table>
Measures and Functions in the Study

**Freshman Data.** The independent variable, Curriculum Type, for the Freshman Data had two levels: SIMMS or non-SIMMS. That is, the subjects of the survey had either taken at least three years of the SIMMS mathematics curriculum in high school or they had not. There were three categories of dependent variables: ACT and SAT scores, freshman mathematics grades, and initial choices of major.

The first variable category of ACT and SAT data had 6 variables: SAT Mathematics, SAT Verbal, ACT Mathematics, ACT Science Reasoning, ACT English, and ACT Reading. These dependent variables took interval scale values. The second variable category also took interval scale values and consisted of freshman mathematics GPA data, disaggregated into three levels: Developmental, Before Calculus, and Calculus. Because the data was disaggregated, there were two independent variables, Curriculum Type and Course Type. The third category of dependent variable was Major Choice, which took nominal scale values, consisting of SMET, non-SMET, or Undeclared, corresponding to the students' initial choices of majors.

**Instructor Survey.** For the college instructor survey there were 12 dependent variables, based on professional standards common to high school and college mathematics practice and corresponding to the questions on the Classroom Practices Inventory (APPENDIX A). Those variables took interval measures, assumed to be on a continuum, on a Likert Scale with selections of "Never," "Seldom," "Sometimes," "Most of
the Time," or "Always." The dependent variables were assumed to be functions of the independent variable, Course Type that had three levels: Developmental, Before Calculus, and Calculus.

Supervisor Interviews. An Interview Protocol (APPENDIX A) was used to guide interviews with eight freshman mathematics course supervisors from University E and eight from University W. One of the supervisors from University W was ill and asked that another veteran instructor who taught his courses be interviewed instead. All freshman courses at both institutions were represented in these interviews. Notes were kept by protocol question and in most cases, audio recordings served as back up devices.

Validity and Reliability in the Study

Freshman Data. The Freshman Data was acquired in the formats of either flat files or Excel™ from the Offices of the Registrars at University E and University W. The data from these files was uploaded to Statistics Package for the Social Sciences (SPSS™), scanned for errors, and processed. A complete record of student data contained the fields listed in Table 3 below.
Table 3. Requested Fields for Freshman Data.

| 1. Student Name                        | 10. Fall 1998 Math Course Number** |
| 2. Student Number                      | 11. Fall 1998 Math Course Grade**   |
| 3. High School Name                    | 12. Spring 1999 Math Course Number**|
| 5. High School GPA                     | 14. Math Placement Exam Score (if available) |
| 6. High School Class Rank              | 15. Math Placement Exam Date (if available) |
| 7. ACT Scores (multiple fields)        | 16. Major declared upon entering     |
| 8. SAT Scores (multiple fields)        |                                           |
| 9. Traditional High School Math vs.    |                                           |
| Integrated Math (SIMMS high school curriculum)* |                                           |

* To qualify for the SIMMS category, a student must have 3 or more Carnegie credits on his/her transcript. If this is not the case, the student defaults into the "Traditional" Category.

** If the student took more than one mathematics course in either semester, the highest course number and its grade are recorded in these fields. If multiple courses were taken and none were passed, the lowest course number and grades (F or incomplete) are recorded. If no courses were taken, leave blank.

Note that the data about majors from both institutions in this study were categorized into one of three Major Choice categories: SMET, non-SMET, or Undeclared categories, according to the schemas in APPENDIX C.

Instructor Survey. The Classroom Practices Inventory (APPENDIX A) was adapted by LaBerge, Zollman, and Sons (1997) from the work of Weiss, Upton, and Nelson (1992) in their NCTM-commissioned study teachers' implementations of the NCTM Standards (LaBerge, Zollman, & Sons, 1997, p. 6). LaBerge, Zollman, and Sons changed a few of the words in their adaptation, presumably to adjust the instrument to the college and university level. They also added a single question to the original eleven contained in the prototype instrument of Weiss et al. That additional question measured
instructors' tendencies to have students "present or discuss solutions to mathematics problems" (LaBerge, Zollman, and Sons, 1997, p. 25).

The Classroom Practice Inventory used in this study was the same as that used by LaBerge, Zollman, and Sons (1997) in their survey of 26 mathematics faculty at seven Midwestern Universities. However, those researchers did not report validity or reliability results for that instrument. Weiss, Matti, and Smith (1994) used all of the questions in that prototype in a large national survey (n=6000) of classroom teachers. Although the latter group reported extensive testing of their instrument (p. 3), they did not give statistics to support validity or reliability. Kull (1996) reported favorable reliability and validity data on an instrument very similar to the one used in this study (p. 136). As an added check in the study at hand, Chronbach's alpha was computed to measure the rationale equivalence reliability (Gay, 1996, p. 149) after data was collected. That statistic provides a lower bound for and, hence, a low estimate of reliability (SPSS base 9.0 applications guide, 1999, p. 362).

**Supervisor Interviews.** After initial construction, the Interview Protocol (APPENDIX A) was reviewed by one mathematics educator and another university educator, the latter, experienced in assessment of higher education. Subsequently, the researcher made changes suggested by both experts.
Research Design

Rationale. This study was essentially three different research pieces. The first piece, or Freshman Data, was descriptive and comparative, and results from it were placed into context via findings from the second and third research pieces. The purpose of the Freshman Data component was to find out if ACT and SAT scores, grades on freshman mathematics courses, and initial choices of major favored SIMMS or non-SIMMS students.

The second research piece, or Instructor Survey, was also descriptive and comparative (Gay, 1996, p. 249). The purpose of the Instructor Survey was to determine if differences in instruction exist among three freshman mathematics Course Types: Developmental, Before Calculus, and Calculus, according to three traditional measures and nine reform-based measures on the Classroom Practices Inventory. A second objective of the Instructor Survey was to distinguish differences, among those Course Types on any of those 12 measures in the Instructor Survey. It should be noted, also, that tests for statistical interactions between the two Curriculum Type and the Course Type, as measured by freshman grades, were considered in the 2 x 3 factorial ANOVA design. Together, then, the two research pieces created snapshots of performance and curricular choices of two potentially different populations of college freshman.

The third research piece was qualitative, consisting of structured interviews with freshman mathematics course supervisors and the collection of relevant course materials (Gay, 1996, p. 215). The purpose of this third component was to determine which goals, objectives, standards, and expectations were built into the freshman mathematics courses.
at University E and University W. Interviews with course supervisors followed a strict protocol, during which supporting documents were solicited. Those documents that alluded to the goals, objectives, standards, or expectations for students were especially considered in the subsequent analysis of data.

Although no pilot studies were done explicitly for either the Freshman Data or the Instructor Survey, previous research by Allinger, Lott, and Lundin (1998), established a need for the work at hand. Included in their report was a framework for the independent variables of this study. The work of Bachelis (1998; 1999) and Milgram (1999) established a need to measure ACT and SAT scores, and the work of LaBerge, Zollman, and Sons (1997) and Kull (1996) established an elegant way to measure classroom practice in an effort to account for potential differences in freshman achievement. Hence, the Classroom Practice Inventory, as previously employed by LaBerge, Zollman, and Sons (1997), was used here without change.

Finally, one desirable effect of a Standards-based curriculum involved pumping students into mathematics and the sciences, rather than filtering them out, as pointed out in the National Science Foundation's Drivers of Reform (MCTM, 1997, p. 2). Therefore, a third variable of the freshman survey, Major Choice, measured SMET, non-SMET, and undeclared majors among incoming freshmen. Consequently, the analysis of data indicated the extent to which former SIMMS students move into the SMET college pipeline when compared with non-SIMMS peers.
Minimization of Invalidity. Several factors potentially affected the validity of this study. For the freshman survey, the ACT and SAT scores, freshman grades, and college major data might not have been what was ordered from sources at either Montana university of interest. To verify validity, the data was scanned and spot-checked against existing data from previous studies (Allinger, Lott, and Lundin, 1998), which had some matching records. Also, non-mathematical SAT and ACT scores were solicited separately from mathematics-related scores at both offices, and duplicates of mathematical data were attached to this second return. Cross-checking the two returns from University E indicated matching numbers of records. However, comparing first and second data sets from University W uncovered some discrepancies. After contacting the registrar, an extra 120 records delivered in that second return from University W were removed from analysis, because they did not represent the requested data profile. The remaining data in that second return matched the profile of the first from Institution W.

The timing of the college Instructor Survey post-dated the collection of the Freshman Data by nearly two years. Thus, it is possible that the snapshot of classroom practices did not accurately reflect the state of classroom practice when freshmen subjects were taking courses. However, it is presumed that the Instructor Survey, because it represents the total respective populations of freshman college instructors at both institutions of interest, did profile freshman instruction for mathematics departments at each Montana university. In keeping with the literature, it was assumed that departmental reform is "usually slow and deliberate" (Stark and Latucca, 1997, p. 372), so that the year's snapshot of classroom practices likely captured a profile of two years prior.
To minimize invalidity that could have resulted from concerns about honest but unpopular responses to questions on the college instructor survey, the letter of introduction to participants guaranteed confidentiality. In addition, department heads at UM and MSU agreed to support the survey, increasing the probability of accurate and timely responses. Finally, to maximize the response rate those participants who had not responded to the survey after two weeks were sent another survey form.

Supervisor interviews were conducted by interview protocol, containing five questions. The format was somewhat strict, although some interviewees were encouraged to say more when the discussion seemed pertinent to the topic at hand. All subjects were recorded, although the recorder malfunctioned for two of these interviews, and for part of a third. In those cases, the researcher's notes and ancillary materials comprised the data.

Procedure

Department heads at both Montana state universities were notified about the college instructor survey and agreed to support its use. Also, registrars at both universities agreed to provide freshman survey data, which was later received. Secondary data sets that contained non-mathematical college entrance exam scores were subsequently solicited and received. The secondary sets were compared to the first with necessary revisions. A third call to each registrar's representative was made to confirm validity of the data.

The researcher sent letters of introduction and survey forms to all freshman course instructors. Those forms were coded by instructor name to check for their return. The resulting packets were bundled with a self-addressed, stamped envelope, in the case of
University W, and an envelope with a campus return address in the case of University E, the researcher's institution. If a completed survey form was not received within two weeks after the mailing, another survey form and letter of introduction was mailed to the instructor.

The researcher entered college instructor survey data into SPSS as it arrived and uploaded freshman performance data into the SPSS software package after it arrived from University E and University W. Files were scanned for proper loading and then spot checked for accuracy against previous data held by the researcher. Routines for coding letter-grade and college major data into numerical data were employed to prepare the data for analysis. The researcher analyzed Freshman Data and data from the college instructor survey in accordance with the methods cited in the Data Analysis Strategy that follows.

Over a period of three weeks, the researcher interviewed eight course supervisors at University W by telephone and eight at University E in person after making appointments with them by e-mail or telephone. All interviews were recorded on two different devices, one, designed for telephone recording. The voice activation failed in two out of the 16 interviews and in one case, the tape ended before the interview did. Structured notes were taken according to protocol question (APPENDIX A), and the researcher solicited materials from the interviewees that pertained to goals, objectives, standards, and objectives for students. These materials were organized and reviewed according to protocol question. Any emerging themes were noted, and a second synthetic analysis was employed to tie findings together.
Data Analysis Strategy

The Freshman Data. Comparing the college entrance examinations of SIMMS and Non-SIMMS freshmen involved six types of measures: SAT Mathematics, SAT Verbal, ACT Mathematics, ACT Science Reasoning, ACT Reading, and ACT English. For each measure an independent t-test was employed to test for significant differences in the mean scores between SIMMS and Non-SIMMS groups. Analysis of data was separated by institution. For each test, $\alpha = .1$. This value was set higher than the standard of .05, because the consequences of not detecting differences in curricular effects (Type I error) were judged to be more severe than detecting them in error (Type II error). That is, a Type I error would result in, at most, a call for more study of the problem, rather than favoritism of one curriculum over another in this case, whereas consequences of a Type II error would mean that performance differences might never be examined. In other words, this researcher wishes to highlight the importance of detecting and examining differences in performance due to curricular influence.

For the analysis of University E college entrance examination data the researcher made the following statistical hypotheses:

1. $H_0$: There is no difference between SAT Mathematics mean scores of former SIMMS and non-SIMMS freshman students at University E in 1998.

2. $H_0$: There is no difference between SAT Verbal mean scores of former SIMMS and non-SIMMS freshman students at University E in 1998.

3. $H_0$: There is no difference between ACT Mathematics mean scores of former SIMMS and non-SIMMS freshman students at University E in 1998.
4. Ho: There is no difference between ACT Science Reasoning mean scores of former SIMMS and non-SIMMS freshman students at University E in 1998.

5. Ho: There is no difference between ACT Reading mean scores of former SIMMS and non-SIMMS freshman students at University E in 1998.

6. Ho: There is no difference between ACT English mean scores of former SIMMS and non-SIMMS freshman students at University E in 1998.

For the analysis of University W college entrance examination data the researcher made the following statistical hypotheses:

1. Ho: There is no difference between SAT Mathematics mean scores of former SIMMS and non-SIMMS freshman students at University W in 1998.

2. Ho: There is no difference between SAT Verbal mean scores of former SIMMS and non-SIMMS freshman students at University W in 1998.

3. Ho: There is no difference between ACT Mathematics mean scores of former SIMMS and non-SIMMS freshman students at University W in 1998.

4. Ho: There is no difference between ACT Science Reasoning mean scores of former SIMMS and non-SIMMS freshman students at University W in 1998.

5. Ho: There is no difference between ACT Reading mean scores of former SIMMS and non-SIMMS freshman students at University W in 1998.

6. Ho: There is no difference between ACT English mean scores of former SIMMS and non-SIMMS freshman students at University W in 1998.

Freshman grades from former SIMMS and non-SIMMS students were analyzed for Fall 1998 and Spring 1999 semesters. Those analyses were performed separately for each institution, using a two-factor ANOVA in all four cases. The dependent variable in
each case was a numerical grade point value (four-point scale), while the two
independent factors were Curriculum Type, with the levels being SIMMS or non-
SIMMS, and Course Type, with levels being Developmental, Before Calculus, and
Calculus. In each case post hoc tests were used to determine which differences among
the levels of the Course Type factor were significant. As in the previous analysis of SAT
and ACT scores, $\alpha = .1$ was chosen, rather than the standard value of .05, because the
researcher believes that the consequences of a Type I error are preferable to those of a
Type II error for the reasons previously stated.

For the analysis of Fall 1998 freshman grades at University E, the researcher
made the following statistical hypotheses related to the mean grade scores of former
SIMMS and non-SIMMS students:

1. $H_0$: There is no significant difference between 1998 Fall mean mathematics grades of
   former SIMMS and non-SIMMS students.

2. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades among the
   three levels of the Course Type factor: Developmental, Before Calculus, or Calculus.

3. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades between
   the Course Type factors, Developmental and Before Calculus.

4. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades between
   the Course Type factors, Developmental and Calculus.

5. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades between
   the Course Type factors, Calculus and Before Calculus.

6. $H_0$: There is no significant interaction in 1998 Fall mean mathematics grades between
   the Course Type and Curriculum Type factors.
For the analysis of Spring 1999 freshman grades at University E, the researcher made the following statistical hypotheses related to the mean grade scores of former SIMMS and non-SIMMS students:

1. $H_0$: There is no significant difference between 1999 Spring mean mathematics grades of former SIMMS and non-SIMMS students at University E.

2. $H_0$: There is no significant difference in 1999 Spring mean mathematics grades at University E among the three levels of the Course Type factor: Developmental, Before Calculus, or Calculus.

3. $H_0$: There is no significant difference in 1999 Spring mean mathematics grades between the Course Type factors, Developmental and Before Calculus.

4. $H_0$: There is no significant difference in 1999 Spring mean mathematics grades between the Course Type factors, Developmental and Calculus.

5. $H_0$: There is no significant difference in 1999 Spring mean mathematics grades between the Course Type factors, Calculus and Before Calculus.

6. $H_0$: There is no significant interaction in 1999 Spring mean mathematics grades between the Course Type and Curriculum Type factors.

For the analysis of Fall 1998 freshman grades at University W, the researcher made the following statistical hypotheses related to the mean grade scores of former SIMMS and non-SIMMS students:

1. $H_0$: There is no significant difference between 1998 Fall mean mathematics grades of former SIMMS and non-SIMMS students.

2. $H_0$: There is no significant difference in 1998 Fall mean mathematics among the three levels of the Course Type factor: Developmental, Before Calculus, or Calculus.
3. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades between the Course Type factors, Developmental and Before Calculus.

4. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades between the Course Type factors, Developmental and Calculus.

5. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades between the Course Type factors, Calculus and Before Calculus.

6. $H_0$: There is no significant interaction in 1998 Fall mean mathematics grades between the Course Type and Curriculum Type factors.

For the analysis of Spring 1999 freshman grades at University W, the researcher made the following statistical hypotheses related to the mean grade scores of former SIMMS and non-SIMMS students:

1. $H_0$: There is no significant difference between 1999 Spring mean mathematics grades of former SIMMS and non-SIMMS students.

2. $H_0$: There is no significant difference in 1999 Spring mean mathematics grades among the three levels of the Course Type factor: Developmental, Before Calculus, or Calculus.

3. $H_0$: There is no significant difference in 1999 Spring mean mathematics grades between the Course Type factors, Developmental and Before Calculus.

4. $H_0$: There is no significant difference in 1999 Spring mean mathematics grades between the Course Type factors, Developmental and Calculus.

5. $H_0$: There is no significant difference in 1999 Spring mean mathematics grades between the Course Type factors, Calculus and Before Calculus.
6. $H_0$: There is no significant interaction in 1999 Spring mean mathematics grades between the Course Type and Curriculum Type factors.

A $\chi^2$ test of independence was employed to determine if initial choice of majors of college freshman was associated with either of two Curriculum Types: SIMMS or non-SIMMS. Curriculum Type is the independent variable, and the dependent variable, Major Choice, had three levels: SMET, non-SMET, or Undeclared. The detailed classifications can be found in APPENDIX C. For each test $\alpha = .1$.

From the analysis of initial choices of majors of 1998 entering Montana freshman at University E and University W, the researcher made the following statistical hypotheses:

1. $H_0$: The proportions of observed counts to expected counts of majors in the categories of SMET, non-SMET, and Undeclared are not significantly different from 1 for former SIMMS and non-SIMMS students at University E.

2. $H_0$: The proportions of observed counts to expected counts of majors in the categories of SMET, non-SMET, and Undeclared are not significantly different from 1 for former SIMMS and non-SIMMS students at University W.

Instructor Survey. For each institution, two new variables, $T_{score}$ and $R_{score}$, were created from the instructor responses to the Classroom Practice Inventory by summing values associated with those responses: one for Never, two for Seldom, three for Sometimes, four for Most of the time, and five for Always. In what follows $R_{score}$ represented the sum of responses to questions two, four, five, seven, eight, ten, eleven, and twelve, associated with classroom reform practices. $T_{score}$ represented the sum of
the numerical responses to the more traditional questions one, three, and nine. Tscore was later discarded as non-representative of the intensity of reform.

A 2 X 3 ANOVA was used to analyze results of Rscores, the Dependent Variable, as a function of two independent variables: Course Type with levels, Developmental, Before Calculus, and Calculus and Institution with levels, University E and University W. For these tests, $\alpha = .05$.

From the analysis of University E and University W Instructor responses to the Survey of Classroom Practices, the researcher made the following statistical hypotheses related to the mean scores of former SIMMS and non-SIMMS students:

1. $H_0$: There is no significant difference between mean Rscores at University E and University W.
2. $H_0$: There is no significant difference among the mean Rscores associated with the Levels of the Curriculum Type Factor: Developmental, Before Calculus, and Calculus.
3. $H_0$: There is no significant difference between mean Rscores for the Before Calculus and Calculus levels of the Curriculum Type factor.
4. $H_0$: There is no significant difference between mean Rscores for the Developmental and Before Calculus levels of the Curriculum Type factor.
5. $H_0$: There is no significant difference between mean Rscores for the Developmental and Calculus levels of the Curriculum Type factor.
6. $H_0$: There are no significant differences among the means associated with interaction between Course Type and Institution Factors.
Supervisor Interviews. Notes from the 16 interviews with course supervisors were reviewed and supplemented as the researcher listened to the interview recordings. Notes were reviewed again to determine emerging themes and representative quotes. These were categorized by Interview Protocol question in matrix form for final write-up.

Assumptions and Limitations of the Study

There were several assumptions concerning the methodology of this study that deserve further attention. It was assumed that data ordered from MSU and UM for the freshman survey was valid. Data were ordered in two distinct batches, which were compared with each other on intersecting data fields to minimize invalidity. In addition, further discussions with both registrars' representatives were needed to account for discrepancies. More information about the validity of the Freshman Data can be found in the next chapter. For the Instructor Survey, it was assumed that a sufficiently large return of the completed survey forms would permit inference to the respective populations. The next chapter also discusses this issue in sufficient detail.

The reader is cautioned about the strict definition of the populations of freshmen students represented in this study. Those populations of students were defined to be 1998 graduates of Montana high schools who enrolled in either University E or University W in Fall 1998. This study makes no inference, nor can it validly do so, outside these parameters. Hence, any inference to 1998 Montana high school graduates who did not attend either University E or University W is neither intended nor implied in this study.

A note on methodology associated with the Freshman Data component of this study is important here. If the object of this study were to construct a linear model for the
prediction of freshman grades, more sophisticated statistical tests would be in order. In particular, a Multivariate Analysis of Variance (MANOVA) could better account for all dependent variables by way of a discriminant function. However, the researcher regarded this as a descriptive study, rather than one whose findings lend themselves to prediction. Indeed, it was the business of this study to be fact-finding rather than predictive. In that spirit, another level of research could be one that uses the power of MANOVA to ascertain which, if any, of the dependent variables are more meaningful than others in their prediction of freshman performance.

Finally, because the size of the SIMMS population is estimated to be one ninth the size of the non-SIMMS population, t-test and ANOVA results may be compromised, in the case of heterogeneous variances. (Glass & Hopkins, 1996, p. 293) In such a case, depending on the ratio of variances of the populations, the probability of a Type I error is affected, and the affect is proportional to both population ratio and variance ratio. Glass and Hopkins (1996) suggest that in this case the Welch's t', a relatively conservative test, be used. (p. 295) In the case of simple ANOVA, a statistical analogue of Welch's test uses non-pooled variance, and this is the statistic of choice for unbalanced designs. (Glass & Hopkins, 1996, p. 405) With multiple factor analysis, such as that required for freshman grades (2 x 3 design), the default option in SPSS uses Type III sums of squares in the computation. This usage is recommended for unbalanced factorial designs. (Glass & Hopkins, 1996, p. 524; SPSS, 1999b, p. 265)
Table 4. Timeframe for the Study.

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 1, 2000</td>
<td>Freshman Survey data sent to Universities E and W</td>
</tr>
<tr>
<td>July 15, 2000</td>
<td>Freshman Data analysis begins</td>
</tr>
<tr>
<td>October 1, 2000</td>
<td>College Instructor Surveys sent to participants at UM and MSU</td>
</tr>
<tr>
<td>October 14, 2000</td>
<td>Instructor Survey reminders sent to participants</td>
</tr>
<tr>
<td>October 1, 2000</td>
<td>Supervisor Interviews Begin</td>
</tr>
<tr>
<td>December 1, 2000</td>
<td>College instructor survey data analysis begins</td>
</tr>
<tr>
<td>January 30, 2001</td>
<td>Data Analysis completed</td>
</tr>
<tr>
<td>January, 15 2001</td>
<td>Writing of Chapters Four and Five begins</td>
</tr>
<tr>
<td>April 23, 2001</td>
<td>Writing of Chapters Four and Five completed</td>
</tr>
</tbody>
</table>

Chapter Summary

In this chapter, the research methodologies for this study were outlined. The Freshman Data, a descriptive and comparative piece, was postulated to detect whether freshman performance favored either the SIMMS curriculum or more traditional high school mathematics curricula. Important measures here included ACT and SAT scores, freshman mathematics grades, and initial choice of majors. Data for this component spanned the populations of former SIMMS students and non-SIMMS students at both University E and University W who graduated from Montana high schools in 1998. T-tests were used to check for significance between groups on ACT and SAT scores. A two-factor ANOVA was used with freshman mathematics grades that were disaggregated by Course Type. For the latter analysis, tests of interactions between independent variables were done, and post-hoc multiple-comparison tests were employed to determine where significant differences lay. A $\chi^2$-test was used to check for significant differences.
among former SIMMS and non-SIMMS groups' initial choices of majors in the three categories: SMET, non-SMET, and Undeclared.

For the classroom Instructor Survey, an ANOVA sorted out potentially significant differences in practices, measured with Rscores, among instructors of three freshman Course Types: Developmental, Before Calculus, and Calculus by Institution. This second descriptive research piece added import and context to the results of the freshman survey.

The Supervisor Interviews with freshman course instructors about their goals, objectives, standards, and expectations for students clarified and enriched findings from the other two components. Together, then, the Freshman Data, Instructor Survey, and the Supervisor Interviews and their respective methodologies comprised this study. Results from these three components are discussed next.
CHAPTER FOUR: RESULTS

Chapter Introduction

Results from this study comparing college-related measures of former SIMMS and non-SIMMS students came via analyses of three data sets from each of the two state universities in Montana. The first set included freshman data, such as ACT and SAT scores, freshman mathematics grades, and initial choices of majors. The second data set consisted of responses to the instructor surveys by freshman mathematics course instructors. The third set of data came from responses to interviews of freshman mathematics course supervisors at both universities.

Generally, data from the two universities were disaggregated by institution, because the two institutions had fundamentally different missions. University E, for example, housed the engineering school, accommodating freshman interested in engineering majors, but University W had no engineering school. Consequently, it was natural to consider those institutions separately when analyzing data, since they most likely attracted different populations of students. Despite a mostly segregated treatment of the two populations, however, a general linear model, encompassing data from both institutions, was used in the analysis of responses to the survey of freshman course instructors. This model permitted a test of interaction between Institution and Course Type variables. Additionally, responses from interviews with freshman course
supervisors about outcomes, standards, and goals for their courses seemed to transcend the division by institution. Therefore, the analysis of the course supervisor responses was not separated by institution except in an ad hoc fashion. Importantly, concern for the anonymity of interviewees dictated that none of their responses identify them. Therefore, descriptions that would do so were deliberately avoided, and this omission included full transcripts of interviews.

Results of the Data Analysis

Freshman Data

Introduction. A representative from the registrar's office at University E returned 1,143 records, representing 81% of the 1,407 students entering that institution with registered Montana addresses in 1998 (Quick Facts for 1998-1999, 1999). The data return rightly did not include the 264 records of students who had graduated Fall 1997, those who had entered that institution by fulfilling GED requirements, and those late transfers into the state (Bob Snyder, Office of the Registrar at University E, personal communication, January 29, 2001). Of those records returned, 1051 represented students who had experience the Non-SIMMS curriculum, while the remainder of 92 records represented former SIMMS students.

A registrar's representative at University W returned 902 freshman records. Importantly, this return represented only 65% of the 1,379 Montana high school graduates who had reportedly entered that institution in 1998. When asked about the difference of 477 records, the registrar's represented commented that these were also due
to late transfers, GED students, and students who did not fit the 1998 graduation profile. These were correctly omitted from the original data query. (C. Burleson, Institutional Research, University W, personal communication on February 5, 2001) Of those 902 records, 802 records represented students who had experienced a non-SIMMS curriculum, and the remaining 100 records represented students from the SIMMS curriculum. A second request for information resulted in a return of 120 more records, which were later determined to be from students who did not fit the 1998 high school graduation qualification of this study.

Classification of the student records into SIMMS or Non-SIMMS categories was determined by matching student names and their documented high school mathematics programs found on transcripts against lists of known SIMMS schools. In cases where a school offered parallel SIMMS and Non-SIMMS curricula, transcripts that reported "Integrated Mathematics," or a similar description of a curriculum, were considered to belong to a former SIMMS student. The number of records attributed to former SIMMS students closely approximated pre-research estimates of 100 students at each institution.

**College Entrance Examinations.** Comparing the college entrance examinations of SIMMS and Non-SIMMS freshmen involved six types of measures: SAT Mathematics, SAT Verbal, ACT Mathematics, ACT Science Reasoning, ACT Reading, and ACT English. For each measure an independent t-test was employed to test for a significant difference in the mean scores between SIMMS and Non-SIMMS groups. Analysis of data was separated by institution. Descriptive statistics, including means and standard deviations for each of the six measures, may be found in Table 5 and Table 6, and the
results of the independent t-tests, including tests for homogeneity of variances, may be found in Table 7 and Table 8. For each test, \( \alpha = .1 \). This value was set higher than the standard of .05, because the consequences of not detecting differences in curricular effects (Type I error) were judged to be more severe than detecting them in error (Type II error). That is, a type I error would result in, at most, a call for more study of the problem, rather than favoritism of one curriculum over another in this case, whereas consequences of a Type II error would mean that performance differences might never be examined. In other words, this researcher wished to highlight the importance of detecting and examining differences in performance due to curricular influence.

Table 5. University E: College Entrance Exam Descriptive Statistics.

<table>
<thead>
<tr>
<th>Curriculum Type</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SATMATH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SIMMS</td>
<td>412</td>
<td>572.14</td>
<td>93.65</td>
</tr>
<tr>
<td>SIMMS</td>
<td>24</td>
<td>501.67</td>
<td>90.49</td>
</tr>
<tr>
<td>SATVERB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SIMMS</td>
<td>412</td>
<td>550.61</td>
<td>91.67</td>
</tr>
<tr>
<td>SIMMS</td>
<td>24</td>
<td>495.42</td>
<td>88.61</td>
</tr>
<tr>
<td>ACTMATH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SIMMS</td>
<td>899</td>
<td>23.76</td>
<td>4.70</td>
</tr>
<tr>
<td>SIMMS</td>
<td>86</td>
<td>21.23</td>
<td>3.73</td>
</tr>
<tr>
<td>ACTSCI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SIMMS</td>
<td>899</td>
<td>23.91</td>
<td>4.34</td>
</tr>
<tr>
<td>SIMMS</td>
<td>86</td>
<td>23.09</td>
<td>3.95</td>
</tr>
<tr>
<td>ACTREAD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SIMMS</td>
<td>899</td>
<td>23.83</td>
<td>5.27</td>
</tr>
<tr>
<td>SIMMS</td>
<td>86</td>
<td>23.07</td>
<td>5.10</td>
</tr>
<tr>
<td>ACTENG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SIMMS</td>
<td>899</td>
<td>22.33</td>
<td>4.54</td>
</tr>
<tr>
<td>SIMMS</td>
<td>86</td>
<td>21.36</td>
<td>4.08</td>
</tr>
</tbody>
</table>
Table 6. University W: College Entrance Exam Descriptive Statistics.

<table>
<thead>
<tr>
<th>Curriculum Type</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SATMATH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SIMMS</td>
<td>278</td>
<td>543.63</td>
<td>77.45</td>
</tr>
<tr>
<td>SIMMS</td>
<td>25</td>
<td>497.60</td>
<td>81.71</td>
</tr>
<tr>
<td>SATVERB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SIMMS</td>
<td>277</td>
<td>543.36</td>
<td>83.75</td>
</tr>
<tr>
<td>SIMMS</td>
<td>25</td>
<td>521.60</td>
<td>76.79</td>
</tr>
<tr>
<td>ACTMATH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SIMMS</td>
<td>534</td>
<td>21.89</td>
<td>4.13</td>
</tr>
<tr>
<td>SIMMS</td>
<td>66</td>
<td>20.47</td>
<td>4.34</td>
</tr>
<tr>
<td>ACTSCI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SIMMS</td>
<td>342</td>
<td>22.56</td>
<td>3.66</td>
</tr>
<tr>
<td>SIMMS</td>
<td>45</td>
<td>22.20</td>
<td>4.52</td>
</tr>
<tr>
<td>ACTREAD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SIMMS</td>
<td>343</td>
<td>23.78</td>
<td>5.24</td>
</tr>
<tr>
<td>SIMMS</td>
<td>45</td>
<td>23.20</td>
<td>4.68</td>
</tr>
<tr>
<td>ACTENG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SIMMS</td>
<td>531</td>
<td>22.04</td>
<td>4.47</td>
</tr>
<tr>
<td>SIMMS</td>
<td>66</td>
<td>21.33</td>
<td>4.63</td>
</tr>
</tbody>
</table>
Table 7. University E: t-values and p-values for College Entrance Examinations.

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>Sig.</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Levene's Test for</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Equality of Variances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>t-test for Equality of Means</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SATMATH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>.001</td>
<td>.972</td>
<td>3.590</td>
<td>434</td>
<td>.000</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>3.701</td>
<td>25.954</td>
<td>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SATVERB</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>.057</td>
<td>.812</td>
<td>2.872</td>
<td>434</td>
<td>.004</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>2.960</td>
<td>25.952</td>
<td>.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ACTMATH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>8.152</td>
<td>.004</td>
<td>4.834</td>
<td>983</td>
<td>.000</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>5.847</td>
<td>112.565</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ACTSCI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>1.188</td>
<td>.276</td>
<td>1.681</td>
<td>983</td>
<td>.093</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>1.818</td>
<td>105.654</td>
<td>.072</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ACTREAD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>.024</td>
<td>.877</td>
<td>1.282</td>
<td>983</td>
<td>.200</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>1.317</td>
<td>103.145</td>
<td>.191</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ACTENG</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>2.203</td>
<td>.138</td>
<td>1.903</td>
<td>983</td>
<td>.057</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>2.080</td>
<td>106.235</td>
<td>.040</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 8. University W: t-values and p-values for College Entrance Examinations.

<table>
<thead>
<tr>
<th></th>
<th>Levene's Test for Equalities of Variances</th>
<th>t-test for Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
</tr>
<tr>
<td>SATMATH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>.180</td>
<td>.672</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>2.709</td>
<td>28.018</td>
</tr>
<tr>
<td>SATVERB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>.035</td>
<td>.852</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>1.346</td>
<td>29.399</td>
</tr>
<tr>
<td>ACTMATH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>.433</td>
<td>.511</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>2.515</td>
<td>80.299</td>
</tr>
<tr>
<td>ACTSCI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>4.185</td>
<td>.041</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>.507</td>
<td>51.901</td>
</tr>
<tr>
<td>ACTREAD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>.475</td>
<td>.491</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>.764</td>
<td>59.437</td>
</tr>
<tr>
<td>ACTENG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>.001</td>
<td>.976</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>1.180</td>
<td>80.820</td>
</tr>
</tbody>
</table>

From the analysis of University E college entrance examination data, the researcher made the following decisions about statistical hypotheses related to the mean scores of former SIMMS and non-SIMMS students. Note that the original statistical hypotheses are reiterated first in each case:
1. $H_0$: There is no difference between SAT Mathematics mean scores of former SIMMS and non-SIMMS freshman students at University E in 1998. For the mean scores on the SAT Mathematics Test, $H_0$ was rejected. The mean SAT Mathematics Test score of former non-SIMMS students was significantly higher than that of former SIMMS students.

2. $H_0$: There is no difference between SAT Verbal mean scores of former SIMMS and non-SIMMS freshman students at University E in 1998. For the mean scores on the SAT Verbal Test, $H_0$ was rejected. The mean SAT Verbal Test score of former non-SIMMS students was significantly higher than that of former SIMMS students.

3. $H_0$: There is no difference between ACT Mathematics mean scores of former SIMMS and non-SIMMS freshman students at University E in 1998. For the mean scores on the ACT Mathematics Test, $H_0$ was rejected. The mean ACT Mathematics Test score of former non-SIMMS students was significantly higher than that of former SIMMS students.

4. $H_0$: There is no difference between ACT Science Reasoning mean scores of former SIMMS and non-SIMMS freshman students at University E in 1998. For the mean scores on the ACT Science Reasoning Test, $H_0$ was rejected. The mean ACT Science Reasoning Test score of former non-SIMMS students was significantly higher than that of former SIMMS students.

5. $H_0$: There is no difference between ACT Reading mean scores of former SIMMS and non-SIMMS freshman students at University E in 1998. For the mean scores on the ACT Reading Test, $H_0$ was retained. The mean ACT Reading Test score of former
non-SIMMS students was not significantly different from that of former SIMMS
students.

6. \( H_0 \): There is no difference between ACT English mean scores of former SIMMS and
non-SIMMS freshman students at University E in 1998. For the mean scores on the
ACT English Test, \( H_0 \) was rejected. The mean ACT English Test score of former non-
SIMMS students was significantly higher than that of former SIMMS students.

From the analysis of University W college entrance examination data the
researcher made the following decisions about statistical hypotheses related to the mean
scores of former SIMMS and non-SIMMS students. Note that the original statistical
hypotheses are reiterated first in each case:

1. \( H_0 \): There is no difference between SAT Mathematics mean scores of former SIMMS
and non-SIMMS freshman students at University W in 1998. For the mean scores on
the SAT Mathematics Test, \( H_0 \) was rejected. The mean SAT Mathematics Test score
of former non-SIMMS students was significantly higher than that of former SIMMS
students.

2. \( H_0 \): There is no difference between SAT Verbal mean scores of former SIMMS and
non-SIMMS freshman students at University W in 1998. For the mean scores on the
SAT Verbal Test, \( H_0 \) was retained. The mean SAT Verbal Test score of former non-
SIMMS students was not significantly different from that of former SIMMS students.

3. \( H_0 \): There is no difference between ACT Mathematics mean scores of former SIMMS
and non-SIMMS freshman students at University W in 1998. For the mean scores on
the ACT Mathematics Test, \( H_0 \) was rejected. The mean ACT Mathematics Test score
of former non-SIMMS students was significantly higher than that of former SIMMS students.

4. $H_0$: There is no difference between ACT Science Reasoning mean scores of former SIMMS and non-SIMMS freshman students at University W in 1998. For the mean scores on the ACT Science Reasoning Test, $H_0$ was retained. The mean ACT Science Reasoning Test score of former non-SIMMS students was not significantly different from that of former SIMMS students.

5. $H_0$: There is no difference between ACT Reading mean scores of former SIMMS and non-SIMMS freshman students at University W in 1998. For the mean scores on the ACT Reading Test, $H_0$ was retained. The mean ACT Reading Test score of former non-SIMMS students was not significantly different from that of former SIMMS students.

6. $H_0$: There is no difference between ACT English mean scores of former SIMMS and non-SIMMS freshman students at University W in 1998. For the mean scores on the ACT English Test, $H_0$ was retained. The mean ACT English score of former non-SIMMS students was not significantly different from that of former SIMMS students.

Summarizing these results, at University E former non-SIMMS students' mean scores significantly differed from those of former SIMMS students on SAT Mathematics and SAT Verbal Tests and on ACT Mathematics, Science Reasoning, and English Tests. At University W only non-SIMMS students' ACT and SAT mathematics mean scores differed significantly from those of former SIMMS students.
Freshman Grades. Freshman grades from former SIMMS and non-SIMMS students were analyzed for Fall 1998 and Spring 1999 semesters. These analyses were performed separately for each institution, using a two factor ANOVA in all four cases. The dependent variable in each case was numerical grade point value (four-point scale), while the two independent factors were Curriculum Type, with the levels being SIMMS or non-SIMMS, and Course Type, with levels being Developmental, Before Calculus, and Calculus. In each case a post hoc Scheffe test was used to determine which differences among the levels of the Course Type factor were significant. As in the previous analysis of SAT and ACT scores, $\alpha = .1$ was chosen, rather than the standard value of .05, because the researcher believes that the consequences of a Type I error are preferable to those of a Type II error for the reasons previously stated. In each of the following four analyses, descriptive statistics were presented in tables to help the narrative flow, while other statistics were presented within the narrative. To aid conceptualization for each of the four analyses, graphs of mean grade points by Course Type and Curriculum Type are also provided.

For Fall 1998 University E grades, a 2 X 3 ANOVA was conducted to evaluate the effects of Curriculum Type and freshman Course Type on freshman mathematics course grades. Means and standard deviations for course grade points (four-point scale) as a function of Curriculum Type and Course Type can be found below in Table 9, and the ANOVA statistics can be found in Table 10. The ANOVA indicated no significant interaction between Course Type and Curriculum Type with $F(2, 679) = .094$, $p = .911$, partial $\eta^2 < .001$. There was, however, a significant difference between SIMMS and non-SIMMS levels of the Curriculum Type variable with $F (1, 679) = 8.970$, $p = .003$, and
partial $\eta^2 = .013$. Likewise, results from the ANOVA indicated a significant difference in the levels of the Course Type variable with $F (2, 679) = 5.595, p = .004$, and $\eta^2 = .016$. A Scheffe Test, used to control for Type I error, indicated a significant difference between mean grades in Developmental courses and those of both Before Calculus and Calculus courses. There was not a significant difference between the mean grades in Before Calculus and Calculus courses. See the graph in Figure 1 below for a visual comparison.

Table 9. Descriptive Statistics for Fall 1998 Grades at University E.

<table>
<thead>
<tr>
<th>Fall Course Code</th>
<th>Curriculum Type</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developmental</td>
<td>Non-SIMMS</td>
<td>2.1189</td>
<td>1.1824</td>
<td>201</td>
</tr>
<tr>
<td></td>
<td>SIMMS</td>
<td>1.4000</td>
<td>1.3540</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.0394</td>
<td>1.2204</td>
<td>226</td>
</tr>
<tr>
<td>Before Calculus</td>
<td>Non-SIMMS</td>
<td>2.9212</td>
<td>1.1276</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>SIMMS</td>
<td>2.3250</td>
<td>1.6998</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.8981</td>
<td>1.1489</td>
<td>103</td>
</tr>
<tr>
<td>Calculus</td>
<td>Non-SIMMS</td>
<td>2.6859</td>
<td>1.2212</td>
<td>334</td>
</tr>
<tr>
<td></td>
<td>SIMMS</td>
<td>1.8455</td>
<td>1.4043</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.6340</td>
<td>1.2477</td>
<td>356</td>
</tr>
<tr>
<td>Total</td>
<td>Non-SIMMS</td>
<td>2.5429</td>
<td>1.2302</td>
<td>634</td>
</tr>
<tr>
<td></td>
<td>SIMMS</td>
<td>1.6647</td>
<td>1.4021</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.4775</td>
<td>1.2639</td>
<td>685</td>
</tr>
</tbody>
</table>
Table 10. ANOVA Table for Fall 1998 Grades at University E.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>97.754(^a)</td>
<td>5</td>
<td>19.551</td>
<td>13.343</td>
<td>.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>500.098</td>
<td>1</td>
<td>500.098</td>
<td>341.294</td>
<td>.000</td>
</tr>
<tr>
<td>CURRIC</td>
<td>13.144</td>
<td>1</td>
<td>13.144</td>
<td>8.970</td>
<td>.003</td>
</tr>
<tr>
<td>F98CSCO</td>
<td>16.397</td>
<td>2</td>
<td>8.198</td>
<td>5.595</td>
<td>.004</td>
</tr>
<tr>
<td>CURRIC * F98CSCO</td>
<td>.274</td>
<td>2</td>
<td>.137</td>
<td>.094</td>
<td>.911</td>
</tr>
<tr>
<td>Error</td>
<td>994.940</td>
<td>679</td>
<td>1.465</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5297.290</td>
<td>685</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>1092.694</td>
<td>684</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) R Squared = .089 (Adjusted R Squared = .083)

Figure 1. Mean Grade Points for Freshman Mathematics Courses at University E in Fall 1998 by Course Type and Curriculum Type.

From the analysis of Fall 1998 freshman grades at University E, the researcher made the following decisions about statistical hypotheses related to the mean scores of
former SIMMS and non-SIMMS students. Note that the original statistical hypotheses are reiterated first in each case:

1. $H_0$: There is no significant difference between 1998 Fall mean mathematics grades of former SIMMS and non-SIMMS students at University E. For the mean scores in freshman grades between former SIMMS and non-SIMMS students, $H_0$ was rejected. The non-SIMMS mean freshman mathematics grade was significantly higher than that of the former SIMMS group.

2. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades at University E among the three levels of the Course Type factor: Developmental, Before Calculus, or Calculus. For the mean scores in freshman grades by Course Type, $H_0$ was rejected. There was a significant difference between at least two of the three means: Developmental, Before Calculus, and Calculus.

3. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades between the Course Type factors, Developmental and Before Calculus. For the mean scores in freshman grades between Developmental and Before Calculus levels, $H_0$ was rejected. The mean mathematics grade score for the Before Calculus level was significantly higher than that of the Developmental Level.

4. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades between the Course Type factors, Developmental and Calculus. For the mean scores in freshman grades between Developmental and Calculus levels, $H_0$ was rejected. The student mean score for the Calculus level was significantly higher than that of the Developmental Level.
5. H₀: There is no significant difference in 1998 Fall mean mathematics grades between the Course Type factors, Calculus and Before Calculus. For the mean scores in freshman grades between Calculus and Before Calculus levels, H₀ was retained. The student mean score for the Calculus level was not significantly different from that of the Before Calculus Level.

6. H₀: There is no significant interaction in 1998 Fall mean mathematics grades between the Course Type and Curriculum Type factors. For each of the differences in means corresponding to the interaction between Course Type and Curriculum Type factors, H₀ was retained. There was no significant interaction between these two factors.

A 2 X 3 ANOVA was again conducted to evaluate the effects of Curriculum Type and freshman Course Type on freshman mathematics course grades. Means and standard deviations for course grade points (four-point scale) as a function of Curriculum Type and Course Type can be found below in Table 11, and the ANOVA statistics, in Table 12. The ANOVA indicated no significant interaction between Course Type and Curriculum Type with F (2, 638) = .453, p = .636, partial η² = .001. There was also no significant difference between SIMMS and non-SIMMS levels of the Curriculum Type variable with F(1, 638) = 2.432, p = .119, and partial η² = .004. However, results from the ANOVA indicated a significant difference in the levels of the Course Type variable with F(1, 638) = 10.827, p < .001, and η² = .033. A Scheffe Test indicated a significant difference between mean grades in Developmental courses and those of both Before Calculus and Calculus courses. There was not a significant difference between mean
grades in the Before Calculus and Calculus courses. See the graph in Figure 2 below for a visual comparison.

Table 11. Descriptive Statistics for Spring 1999 Grades at University E.

<table>
<thead>
<tr>
<th>Curriculum Type · Spring 99 Course Code</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-SIMMS Developmental</td>
<td>1.5485</td>
<td>1.2063</td>
<td>136</td>
</tr>
<tr>
<td>Before Calculus</td>
<td>2.6248</td>
<td>1.1410</td>
<td>145</td>
</tr>
<tr>
<td>Calculus</td>
<td>2.4694</td>
<td>1.2452</td>
<td>310</td>
</tr>
<tr>
<td>Total</td>
<td>2.2956</td>
<td>1.2783</td>
<td>591</td>
</tr>
<tr>
<td>SIMMS Developmental</td>
<td>1.4500</td>
<td>1.3169</td>
<td>20</td>
</tr>
<tr>
<td>Before Calculus</td>
<td>2.0000</td>
<td>1.5373</td>
<td>7</td>
</tr>
<tr>
<td>Calculus</td>
<td>2.2462</td>
<td>1.1752</td>
<td>26</td>
</tr>
<tr>
<td>Total</td>
<td>1.9132</td>
<td>1.3074</td>
<td>53</td>
</tr>
<tr>
<td>Total • Developmental</td>
<td>1.5359</td>
<td>1.2170</td>
<td>156</td>
</tr>
<tr>
<td>Before Calculus</td>
<td>2.5961</td>
<td>1.1631</td>
<td>152</td>
</tr>
<tr>
<td>Calculus</td>
<td>2.4521</td>
<td>1.2397</td>
<td>336</td>
</tr>
<tr>
<td>Total</td>
<td>2.2641</td>
<td>1.2840</td>
<td>644</td>
</tr>
</tbody>
</table>

Table 12. ANOVA Table for Fall 1998 Grades at University E.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>115.318(^a)</td>
<td>5</td>
<td>23.064</td>
<td>15.575</td>
<td>.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>611.942</td>
<td>1</td>
<td>611.942</td>
<td>413.263</td>
<td>.000</td>
</tr>
<tr>
<td>CURRIC</td>
<td>3.601</td>
<td>1</td>
<td>3.601</td>
<td>2.432</td>
<td>.119</td>
</tr>
<tr>
<td>S99CSCD</td>
<td>32.063</td>
<td>2</td>
<td>16.031</td>
<td>10.827</td>
<td>.000</td>
</tr>
<tr>
<td>CURRIC * S99CSCD</td>
<td>1.343</td>
<td>2</td>
<td>.671</td>
<td>.453</td>
<td>.636</td>
</tr>
<tr>
<td>Error</td>
<td>944.724</td>
<td>638</td>
<td>1.481</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4361.370</td>
<td>644</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>1060.041</td>
<td>643</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) R Squared = .109 (Adjusted R Squared = .102)
Figure 2. Mean Grade Points for Freshman Mathematics Courses at University E in Spring 1999 by Course Type and Curriculum Type.

From the analysis of Spring 1999 freshman grades at University E, the researcher made the following decisions about statistical hypotheses related to the mean scores of former SIMMS and non-SIMMS students. Note that the original statistical hypotheses are reiterated first in each case:

1. $H_0$: There is no significant difference between 1999 Spring mean mathematics grades of former SIMMS and non-SIMMS students at University E. For the mean scores in freshman grades between former SIMMS and non-SIMMS students, $H_0$ was retained. The Non-SIMMS mean freshman mathematics grade was not significantly different from that of the former SIMMS group.
2. $H_0$: There is no significant difference in 1999 Spring mean mathematics grades at University E among the three levels of the Course Type factor: Developmental, Before Calculus, or Calculus. For the mean scores in freshman grades by Course Type, $H_0$ was rejected. There was a significant difference between at least two of the three means: Developmental, Before Calculus, and Calculus.

3. $H_0$: There is no significant difference in 1999 Spring mean mathematics grades at University E between the Course Type factors, Developmental and Before Calculus. For the mean scores in freshman grades between Developmental and Before Calculus levels, $H_0$ was rejected. The student mean score for the Before Calculus level was significantly higher than that of the Developmental Level.

4. $H_0$: There is no significant difference in 1999 Spring mean mathematics grades at University E between the Course Type factors, Developmental and Calculus. For the mean scores in freshman grades between Developmental and Calculus levels, $H_0$ was rejected. The student mean score for the Calculus level was significantly higher than that of the Developmental Level.

5. $H_0$: There is no significant difference in 1999 Spring mean mathematics grades at University E between the Course Type factors, Calculus and Before Calculus. For the mean scores in freshman grades between Calculus and Before Calculus levels, $H_0$ was retained. The student mean score for the Calculus level was not significantly different from that of the Before Calculus Level.

6. $H_0$: There is no significant interaction in 1999 Spring mean mathematics grades at University E between the Course Type and Curriculum Type factors. For each of the differences in means corresponding to the interaction between Course Type and
Curriculum Type factors, $H_0$ was retained. There was no significant interaction between these two factors.

A $2 \times 3$ ANOVA was conducted to evaluate the effects of Curriculum Type and freshman Course Type on freshman mathematics course grades. Means and standard deviations for course grade points (four-point scale) as a function of Curriculum Type and Course Type can be found below in Table 13, and the ANOVA statistics, in Table 14. The ANOVA indicated no significant interaction between Course Type and Curriculum Type with $F(2, 568) = .195$, $p = .823$, partial $\eta^2 = .001$. Also, there was no significant difference between SIMMS and non-SIMMS levels of the Curriculum Type variable with $F(1, 568) = 1.427$, $p = .233$, and partial $\eta^2 = .003$. However, results from the ANOVA indicated a significant difference in the levels of the Course Type variable with $F(2, 568) = 7.825$, $p = .001$, and $\eta^2 = .027$. A Scheffe Test indicated a significant difference between mean grades in Developmental courses and those of both Before Calculus and Calculus courses. There was not a significant difference between mean grades in the Before Calculus and Calculus courses. See the graph in Figure 3 below for a visual comparison.
Table 13. Descriptive Statistics for Fall 1998 Grades at University W.

<table>
<thead>
<tr>
<th>Curriculum Type</th>
<th>Fall 98 Course Type</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-SIMMS</td>
<td>Developmental</td>
<td>2.03</td>
<td>1.30</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>Before Calculus</td>
<td>2.67</td>
<td>1.23</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td>Calculus</td>
<td>2.60</td>
<td>1.26</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.48</td>
<td>1.28</td>
<td>515</td>
</tr>
<tr>
<td>SIMMS</td>
<td>Developmental</td>
<td>1.67</td>
<td>1.27</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Before Calculus</td>
<td>2.39</td>
<td>1.34</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Calculus</td>
<td>2.56</td>
<td>1.13</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.08</td>
<td>1.32</td>
<td>59</td>
</tr>
<tr>
<td>Total</td>
<td>Developmental</td>
<td>1.97</td>
<td>1.30</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>Before Calculus</td>
<td>2.65</td>
<td>1.24</td>
<td>249</td>
</tr>
<tr>
<td></td>
<td>Calculus</td>
<td>2.60</td>
<td>1.25</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.44</td>
<td>1.29</td>
<td>574</td>
</tr>
</tbody>
</table>

Table 14. ANOVA Table for Fall 1998 Grades at University W.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>55.454$^a$</td>
<td>5</td>
<td>11.091</td>
<td>6.984</td>
<td>.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>922.538</td>
<td>1</td>
<td>922.538</td>
<td>580.912</td>
<td>.000</td>
</tr>
<tr>
<td>CTYPE</td>
<td>2.266</td>
<td>1</td>
<td>2.266</td>
<td>1.427</td>
<td>.233</td>
</tr>
<tr>
<td>FCRSCOD</td>
<td>24.853</td>
<td>2</td>
<td>12.427</td>
<td>7.825</td>
<td>.000</td>
</tr>
<tr>
<td>CTYPE * FCRSCOD</td>
<td>.620</td>
<td>2</td>
<td>.310</td>
<td>.195</td>
<td>.823</td>
</tr>
<tr>
<td>Error</td>
<td>902.032</td>
<td>568</td>
<td>1.588</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4377.000</td>
<td>574</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ R Squared = .058 (Adjusted R Squared = .050)
From the analysis of Fall 1998 freshman grades at University W, the researcher made the following decisions about statistical hypotheses related to the mean scores of former SIMMS and non-SIMMS students. Note that the original statistical hypotheses are reiterated first in each case:

1. $H_0$: There is no significant difference between 1998 Fall mean mathematics grades of former SIMMS and non-SIMMS students at University W. For the mean scores in freshman grades between former SIMMS and non-SIMMS students, $H_0$ was retained. The Non-SIMMS mean freshman mathematics grade was not significantly different from that of the former SIMMS group.

Figure 3. Mean Grade Points for Freshman Mathematics Courses at University W in Fall 1998 by Course Type and Curriculum Type.
2. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades at University W among the three levels of the Course Type factor: Developmental, Before Calculus, or Calculus. For the mean scores in freshman grades by Course Type, $H_0$ was rejected. There was a significant difference between at least two of the three means: Developmental, Before Calculus, and Calculus.

3. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades at University W between the Course Type factors, Developmental and Before Calculus. For the mean scores in freshman grades between Developmental and Before Calculus levels, $H_0$ was rejected. The mean score for the Before Calculus level was significantly higher than that of the Developmental Level.

4. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades at University W between the Course Type factors, Developmental and Calculus. For the mean scores in freshman grades between Developmental and Calculus levels, $H_0$ was rejected. The mean score for the Calculus level was significantly higher than that of the Developmental Level.

5. $H_0$: There is no significant difference in 1998 Fall mean mathematics grades at University W between the Course Type factors, Calculus and Before Calculus. For the mean scores in freshman grades between Calculus and Before Calculus levels, $H_0$ was retained. The mean score for the Calculus level was not significantly different from that of the Before Calculus Level.

6. $H_0$: There is no significant interaction in 1998 Fall mean mathematics grades at University W between the Course Type and Curriculum Type factors. For each of the differences in means corresponding to the interaction between Course Type and
Curriculum Type factors, $H_0$ was retained. There was no significant interaction between these two factors.

As in the Fall 1998 case, a 2 X 3 ANOVA was conducted to evaluate the effects of Curriculum Type and freshman Course Type on freshman mathematics course grades. Means and standard deviations for course grade points (four-point scale) as a function of Curriculum Type and Course Type can be found below in Table 15, and the ANOVA statistics, in Table 16. The ANOVA indicated no significant interaction between Course Type and Curriculum Type with $F(2, 397) = 1.378$, $p = .253$, partial $\eta^2 = .007$.

However, there was a significant difference between SIMMS and non-SIMMS levels of the Curriculum Type variable with $F(1, 397) = 3.480$, $p = .063$, and partial $\eta^2 = .009$. Also, results from the ANOVA indicated a significant difference in the levels of the Course Type variable with $F(2, 397) = 16.504$, $p < .001$, and $\eta^2 = .077$. As for the previous semester's analysis at this institution, a Scheffe Test indicated a significant difference between mean grades in Developmental courses and those of both Before Calculus and Calculus courses. There was not a significant difference between the mean grades of Before Calculus and Calculus courses. See the graph in Figure 4 below for a visual comparison.
Table 15. Descriptive Statistics for Spring 1999 Grades at University W.

<table>
<thead>
<tr>
<th>Curriculum Type</th>
<th>Spring 99 Course Type</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-SIMMS</td>
<td>Developmental</td>
<td>1.57</td>
<td>1.07</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Before Calculus</td>
<td>2.60</td>
<td>1.22</td>
<td>220</td>
</tr>
<tr>
<td></td>
<td>Calculus</td>
<td>2.75</td>
<td>1.18</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.53</td>
<td>1.24</td>
<td>359</td>
</tr>
<tr>
<td>SIMMS</td>
<td>Developmental</td>
<td>.57</td>
<td>.79</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Before Calculus</td>
<td>2.50</td>
<td>1.32</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Calculus</td>
<td>2.62</td>
<td>1.12</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.23</td>
<td>1.38</td>
<td>44</td>
</tr>
<tr>
<td>Total</td>
<td>Developmental</td>
<td>1.41</td>
<td>1.09</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Before Calculus</td>
<td>2.59</td>
<td>1.23</td>
<td>244</td>
</tr>
<tr>
<td></td>
<td>Calculus</td>
<td>2.73</td>
<td>1.17</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.50</td>
<td>1.25</td>
<td>403</td>
</tr>
</tbody>
</table>

Table 16. ANOVA Table for Spring 1999 Grades at University W.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>66.705a</td>
<td>5</td>
<td>13.341</td>
<td>9.357</td>
<td>.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>524.223</td>
<td>1</td>
<td>524.223</td>
<td>367.668</td>
<td>.000</td>
</tr>
<tr>
<td>CTYPE</td>
<td>4.962</td>
<td>1</td>
<td>4.962</td>
<td>3.480</td>
<td>.063</td>
</tr>
<tr>
<td>SCRSCOD</td>
<td>47.063</td>
<td>2</td>
<td>23.531</td>
<td>16.504</td>
<td>.000</td>
</tr>
<tr>
<td>CTYPE * SCRSCOD</td>
<td>3.929</td>
<td>2</td>
<td>1.965</td>
<td>1.378</td>
<td>.253</td>
</tr>
<tr>
<td>Error</td>
<td>566.045</td>
<td>397</td>
<td>1.426</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3154.000</td>
<td>403</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>632.749</td>
<td>402</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. R Squared = .105 (Adjusted R Squared = .094)
From the analysis of Spring 1999 freshman grades at University W, the researcher made the following decisions about statistical hypotheses related to the mean scores of former SIMMS and non-SIMMS students. Note that the original statistical hypotheses are reiterated first in each case:

1. $H_0$: There is no significant difference between Spring 1999 mean mathematics grades of former SIMMS and non-SIMMS students at University W. For the mean scores in freshman grades between former SIMMS and non-SIMMS students, $H_0$ was rejected. The Non-SIMMS mean freshman mathematics grade was significantly higher than that of the former SIMMS group.
2. $H_0$: There is no significant difference in Spring 1999 mean mathematics grades at University W among the three levels of the Course Type factor: Developmental, Before Calculus, or Calculus. For the mean scores in freshman grades by Course Type, $H_0$ was rejected. There was a significant difference between at least two of the three means: Developmental, Before Calculus, and Calculus.

3. $H_0$: There is no significant difference in Spring 1999 mean mathematics grades at University W between the Course Type factors, Developmental and Before Calculus. For the mean scores in freshman grades between Developmental and Before Calculus levels, $H_0$ was rejected. The student mean score for the Before Calculus level was significantly higher than that of the Developmental Level.

4. $H_0$: There is no significant difference in 1999 Spring mean mathematics grades at University W between the Course Type factors, Developmental and Calculus. For the mean scores in freshman grades between Developmental and Calculus levels, $H_0$ was rejected. The student mean score for the Calculus level was significantly higher than that of the Developmental Level.

5. $H_0$: There is no significant difference in Spring 1999 mean mathematics grades at University W between the Course Type factors, Calculus and Before Calculus. For the mean scores in freshman grades between Calculus and Before Calculus levels, $H_0$ was retained. The student mean score for the Calculus level was not significantly different from that of the Before Calculus Level.

6. $H_0$: There is no significant interaction in Spring 1999 mean mathematics grades at University W between the Course Type and Curriculum Type factors. For each of the differences in means corresponding to the interaction between Course Type and
Curriculum Type factors, $H_0$ was retained. There was no significant interaction between these two factors.

**Choice of Majors.** A $\chi^2$ test of independence was employed to determine if initial choice of majors of college freshman was associated with either of two curriculum types: SIMMS or non-SIMMS. Hence, Curriculum Type is the independent variable here. The dependent variable, Major Choice, has three levels: SMET, non-SMET, or Undeclared. Majors were categorized, when in doubt of their proper taxonomy, by their descriptions in respective institutional catalogs. The detailed classifications can be found in APPENDIX C. For each test $\alpha = .1$.

For University E a $\chi^2$ test of independence indicated no association of Major Choice with Curriculum Type, despite a slightly higher percentage (2%) of choices of SMET majors for the SIMMS group. For this case $\chi^2 (2, N = 1142) = .810$ for $p = .667$. For University W the $\chi^2$ test for independence also indicated no association of Major Choice with Curriculum Type, again with a 2% higher choice of SMET majors among the SIMMS group. For this case $\chi^2 (2, N = 902) = 1.08$ with $p = .583$. Crosstabulation tables with the results for both institutions can be found in Table 17 and Table 18 below.
Table 17. University E: Curriculum Type vs. Major Choice of 1998 Freshmen.

<table>
<thead>
<tr>
<th>Non-SIMMS</th>
<th>Count</th>
<th>SMET</th>
<th>Non-SMET</th>
<th>Undeclared</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>477.</td>
<td>276</td>
<td>297</td>
<td>1050</td>
</tr>
<tr>
<td>SIMMS</td>
<td>Expected Count</td>
<td>479.0</td>
<td>277.7</td>
<td>293.3</td>
<td>1050.0</td>
</tr>
<tr>
<td></td>
<td>% within Curriculum Type</td>
<td>45.4%</td>
<td>26.3%</td>
<td>28.3%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Total</td>
<td>Count</td>
<td>521</td>
<td>302</td>
<td>319</td>
<td>1142</td>
</tr>
<tr>
<td></td>
<td>Expected Count</td>
<td>521.0</td>
<td>302.0</td>
<td>319.0</td>
<td>1142.0</td>
</tr>
<tr>
<td></td>
<td>% within Curriculum Type</td>
<td>45.6%</td>
<td>26.4%</td>
<td>27.9%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 18. University W: Curriculum Type vs. Major Choice of 1998 Freshmen.

<table>
<thead>
<tr>
<th>Non-SIMMS</th>
<th>Count</th>
<th>SMET</th>
<th>Non-SMET</th>
<th>Undeclared</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>152</td>
<td>389</td>
<td>261</td>
<td>802</td>
</tr>
<tr>
<td>SIMMS</td>
<td>Expected Count</td>
<td>153.8</td>
<td>384.1</td>
<td>264.1</td>
<td>802.0</td>
</tr>
<tr>
<td></td>
<td>% within Curriculum Type</td>
<td>19.0%</td>
<td>48.5%</td>
<td>32.5%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Total</td>
<td>Count</td>
<td>173</td>
<td>432</td>
<td>297</td>
<td>902</td>
</tr>
<tr>
<td></td>
<td>Expected Count</td>
<td>173.0</td>
<td>432.0</td>
<td>297.0</td>
<td>902.0</td>
</tr>
<tr>
<td></td>
<td>% within Curriculum Type</td>
<td>19.2%</td>
<td>47.9%</td>
<td>32.9%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

From the analysis of initial choices of majors of 1998 entering Montana freshman at University E and University W, the researcher made the following decisions about
statistical hypotheses. Note that the original statistical hypotheses are reiterated first in each case:

1. \( H_0 \): The proportions of observed counts to expected counts of majors in the categories of SMET, non-SMET, and Undeclared are not significantly different from 1 for former SIMMS and non-SIMMS students at University E. \( H_0 \) was retained. At University E, the proportions of observed counts to expected counts of majors in the categories of SMET, non-SMET, and Undeclared were not significantly different from 1 for former SIMMS and non-SIMMS students.

2. \( H_0 \): The proportions of observed counts to expected counts of majors in the categories of SMET, non-SMET, and Undeclared are not significantly different from 1 for former SIMMS and non-SIMMS students at University W. \( H_0 \) was retained. At University W, the proportions of observed counts to expected counts of majors in the categories of SMET, non-SMET, and Undeclared were not significantly different from 1 for former SIMMS and non-SIMMS students.

**Instructor Surveys**

**Introduction.** Return rates for the completed survey forms were relatively high, 63 of 70 (90%) from University E, but only moderately high, 58 of 74 (78%), from University W. The high return rate from University E permits inference of the subsequent findings to each of the Course Type subpopulations (Gay, 1996, p.125). From that institution, nine of ten (90%) surveys were returned from Developmental Instructors, 25 of 26 (96%), from Course Before Calculus instructors, and 31 of 34 (91%) were returned from Calculus instructors.
In contrast, the rate of return from University W, although moderately high, was insufficient to permit inferences to any of the three subpopulations, except marginally, at the Developmental level (Gay, 1996, p. 25). From University W, ten of twelve (83%) Developmental instructors responded along with 26 of 33 (79%) of Course Before Calculus Instructors, and 21 of 29 (72%) of Calculus Instructors.

**Rscore Analysis.** For each institution, two new variables, Tscore and Rscore, were created from the instructor responses by summing values associated with those responses: one for "Never," two for "Seldom," three for "Sometimes," four for "Most of the Time," and five for "Always." In what follows Rscore represents the sum of responses to questions two, four, five, seven, eight, ten, eleven, and twelve associated with classroom reform practices. Tscore represents the sum of the numerical responses to the more traditional questions one, three, and nine.

With respect to each institution, a bivariate linear regression produced significant positive correlation between the traditional and reform variables, Tscore and Rscore, rather than the expected negative correlation. For University E, that \( r = .453 \) with \( t (61) = 3.97, p < .0001 \). For University W, there was a lesser correlation value of \( r = .312 \) with \( t (55) = 2.43, p = .018 \). In both analyses Tscore was considered to be the independent variable, explaining 20% and 10% of the variance of Rscore, respectively, at University E and University W. Tscore was exempted from further analysis, since it did not, as expected, measure practices contrasting to those of reform.

Rational equivalence reliability, a measure of internal consistency of the survey instrument (Gay, 1996, p. 149), was computed in two ways using the method of
Cronbach. The calculation of Cronbach's $\alpha$ required that all questions measuring the same construct be of the same polarity (Green, Salkind & Akey, 2000, p. 315). Because of the positive correlation of the reform variable, Rscore, to the traditional variable, Tscore, Cronbach's $\alpha$ was computed with and without reversing the order of the values of Tscore. Without transforming the values of the responses to the traditional questions one, three, and nine, Cronbach's $\alpha = .7632$ for $n = 120$ from the pooled University E and University W data. After reversing values on those three traditional questions, $\alpha = .6151$, lower than the previous value, which was reasonable, since Rscore was positively correlated to Tscore for both sets of data.

Results from a 2 X 3 ANOVA showed significant mean Rscore differences between University E and University W and also between instructors teaching calculus and instructors teaching courses before calculus. There were, however, no significant differences between either Calculus or Before Calculus mean Rscores and those of Developmental course instructors. For the institutional differences, $F (1, 114) = 14.8$, $p < .001$ with $\eta^2 = .12$, and for the differences in Course Type, $F (2, 114) = 4.56$, $p < .001$, and $\eta^2 = .74$. There was no significant interaction between Institution and Course Type: $F (2, 114) = 1.32$, $p = .271$, and $\eta^2 = .02$. A Scheffe post hoc procedure was used to control for Type I error that might otherwise result from pairwise comparisons of the levels of the Course Type factor. The results of the Scheffe analysis indicate a significant difference ($p = .003$) between the Calculus and Before Calculus levels but no significant differences between other levels of that factor. A table of descriptive statistics, including means and standard deviations, and a comparative graph of these results may be found.
below, respectively, in Table 19 and Figure 5. The ANOVA table of statistics can be found below in Table 20.

Table 19. Descriptive Statistics for Mean Responses on Instructor Surveys.

<table>
<thead>
<tr>
<th>Institution</th>
<th>Course Type</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>University E</td>
<td>Developmental</td>
<td>24.0000</td>
<td>2.0000</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Course Before Calculus</td>
<td>27.0400</td>
<td>6.2081</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Calculus</td>
<td>22.5000</td>
<td>3.6742</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>24.4921</td>
<td>5.1239</td>
<td>63</td>
</tr>
<tr>
<td>University W</td>
<td>Developmental</td>
<td>28.5000</td>
<td>5.4006</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Course Before Calculus</td>
<td>29.1154</td>
<td>5.3764</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Calculus</td>
<td>27.7143</td>
<td>4.8077</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>28.4912</td>
<td>5.1239</td>
<td>57</td>
</tr>
<tr>
<td>Total</td>
<td>Developmental</td>
<td>26.5000</td>
<td>4.7310</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Course Before Calculus</td>
<td>28.0980</td>
<td>5.8353</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Calculus</td>
<td>24.6471</td>
<td>4.8778</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>26.3917</td>
<td>5.4823</td>
<td>120</td>
</tr>
</tbody>
</table>

Table 20. ANOVA Statistics Comparing Rscore by Institution and Course Type.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>784.692&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5</td>
<td>156.938</td>
<td>6.408</td>
<td>.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>65657.280</td>
<td>1</td>
<td>65657.280</td>
<td>2680.945</td>
<td>.000</td>
</tr>
<tr>
<td>SCHOOL</td>
<td>361.580</td>
<td>1</td>
<td>361.580</td>
<td>14.764</td>
<td>.000</td>
</tr>
<tr>
<td>CTYPE</td>
<td>223.434</td>
<td>2</td>
<td>111.717</td>
<td>4.562</td>
<td>.012</td>
</tr>
<tr>
<td>SCHOOL * CTYPE</td>
<td>64.728</td>
<td>2</td>
<td>32.364</td>
<td>1.321</td>
<td>.271</td>
</tr>
<tr>
<td>Error</td>
<td>2791.900</td>
<td>114</td>
<td>24.490</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>87159.000</td>
<td>120</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>3576.592</td>
<td>119</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>: R Squared = .219 (Adjusted R Squared = .185)
From the analysis of University E and University W instructor responses to the Instructor Survey, the researcher made the following decisions about statistical hypotheses related to the mean scores of former SIMMS and non-SIMMS students. Note that the original statistical hypotheses are reiterated first in each case:

1. $H_0$: There is no significant difference between mean Rscores at University E and University W. $H_0$ was rejected. The University W mean Rscore was significantly higher than the University E mean Rscore.

2. $H_0$: There is no significant difference among the mean Rscores associated with the Levels of the Curriculum Type Factor: Developmental, Before Calculus, and Calculus. $H_0$ was rejected. At least two of the mean Rscores of Developmental, Before Calculus, and Calculus levels were significantly different.
3. $H_0$: There is no significant difference between mean Rscores for the Before Calculus and Calculus levels of the Curriculum Type factor. $H_0$ was rejected. The mean Rscore of the Before Calculus groups was significantly higher than that of the Calculus Group.

4. $H_0$: There is no significant difference between mean Rscores for the Developmental and Before Calculus levels of the Curriculum Type factor. $H_0$ was retained. No significant difference existed between mean Rscore responses from Developmental and Before Calculus instructors.

5. $H_0$: There is no significant difference between mean Rscores for the Developmental and Calculus levels of the Curriculum Type factor. $H_0$ was retained. No significant difference existed between mean Rscore responses from Developmental and Calculus instructors.

6. $H_0$: There are no significant differences among of the means associated with interaction between Course Type and Institution Factors. $H_0$ was retained. There was no significant interaction between Course Type and Institution.

**Analysis of Classroom Practices by Survey Question.** Responses for each of the twelve questions of the Classroom Practices Inventory were averaged separately for University E and University W, and those averages were plotted by Course Type in Figure 6 and Figure 7. The graphic profiles for both Universities shared a similar in "topography," with peaks and valleys in group mean responses centered on the same questions. Peak mean response maximums and valley minimums slightly differed by institution for certain questions, however, with calculus mean responses at University W showing the most divergence from a general profile.
Figure 6. University E: Instructor Response Means by Course Type.

Figure 7. University W: Instructor Response Means by Course Type.
Notably, instructors from University W rated themselves higher on the average than did instructors from University E, and their responses were also more varied. The mean response value from instructors at University E was 2.47 with a variance of .778, while the mean response value from instructors at University W was 3.76 with a variance of 4.48. These findings paralleled those cited previously with Rscores, but the line graphs bring visual clarity to patterns of responses.

Again, considering these graphs, it was clear that the reform profiles at each institution were similar, despite the overall higher responses of University W. Because the "topography" of the two graphs was the same, analysis of the results allows natural categorization of classroom practices into three levels: More Frequent Practices, Less Frequent Practices and Moderately Frequent Practices. More Frequent Practices at both institutions were those whose values on both graphs tended to peak, while Less Frequent Practices had values in troughs. Moderately Frequent Practices had values lying between peaks and troughs.

Clearly from the two graphs, More Frequent Practices in the freshman mathematics classroom included 1) using problems from the text, 2) employing real-world problems, 3) using calculators, and 4) taking notes. Less Frequent Practices include 1) using physical models, 2) facilitating cooperative learning, 3) assigning in-class projects, and 4) using computers in class. Moderately Frequent Practices include 1) using worksheets, 2) incorporating problem solving, 3) writing about mathematics, and 4) Presenting solutions to problems.

On several of the survey items, group means "broke" profile. In particular, on the average, calculus instructors at University E rated themselves lower in using group work,
and the Developmental instructors at the same institution rated themselves higher than the other two groups in using calculators. At University W, the Calculus group had self-ratings 1.5 points below the Developmental and Before Calculus Groups in note taking in the classroom and a half-point higher in using computers.

Supervisor Interviews

Introduction. The researcher conducted 16 interviews with freshman mathematics course supervisors, eight from University E and eight from University W. Interviewees included the supervisor of developmental courses at each institution, four supervisors of Before Calculus courses at University E and five at University W, and three supervisors of Calculus courses (including precalculus) at University E and two at University W. Every freshman mathematics course in those three categories at either institution was supervised by one of the interviewees, except for one. In that case the supervisor was ill and suggested that the researcher speak to one of his instructors who had taken on some of his duties. Interviews lasted from 20 minutes to over an hour, and all but two were taped. All interviews at University W were conducted by telephone, while those at University E were conducted in person. The researcher took notes following a protocol of five interview questions. (See the Interview Protocol in APPENDIX A.) The interview protocol focused on the topic of assessment of goals, outcomes, standards, and expectations for students. Because of this rather rigid outline, note taking by protocol question lent itself to analysis of data, and the tapes were used as back-up resources. The rigidity of the interview protocol made transcription of tapes superfluous, and, importantly, several of the interviewees expressed their desire for anonymity. Therefore,
transcriptions of tapes were not produced. Regardless of the relative rigidity of the interview protocol, some interviewees were encouraged to digress when the topics seemed pertinent to the research at hand. Reviews of notes and tapes resulted in several themes implicitly related to the research in addition to those themes explicit in the protocol questions.

The researcher analyzed responses to the interview protocol by question and by implicit themes. To aid the analysis, the researcher constructed a matrix of responses, coded by question and course supervisor. To smooth the narrative, implicit themes were considered within the analysis of each question, and this is the subject of the next section.

Analysis of the Responses of Supervisor Interviews.

1. Do you have a written set of goals, outcomes, standards, or expectations against which students' achievement may be measured?

Of the 16 course supervisors who were interviewed, 14 shared syllabi and other materials with the researcher or alluded to web sites containing those materials: www.math.umt.edu.classes/ at University W and www.math.montana.edu/courses/courses.html at University E. Six courses relevant to this study from University E and six from University W were represented at the respective web sites.

Common to all materials were schedules, course content outlines, and grading schemes. Many of the syllabi referred to chapters in texts. In three cases, materials included suggestions for learning, studying, or doing mathematics. In only one case were goals specifically stated, which included ten expectations for students that involved
reading, writing, and doing mathematics (Esty, 2001). In another case the course coordinator mentioned that goals were specifically stated in the texts for the (Developmental) courses.

Materials from the Developmental course instructors at each institution contained procedural guides, including, at University W, a quiz on course procedures to be completed by students. Implicit in all of these materials, evidently, was the importance of surviving the content of the courses, given the limits and guidelines set forth by the syllabi. However, just because these materials lacked explicit goals, expectations, outcomes, or standards does not mean that these are never communicated to instructors or students. Indeed, one veteran calculus instructor who is not a course supervisor posted clear course goals along with her syllabus (McNulty, J., 2000). More generally, though, these syllabi strongly imply that meeting course goals meant learning the content according to some schedule, the content, outlined therein.

2. What goals, outcomes, standards, or expectations for the students are stressed by you and your instructors? (In other words, what do you want your students to be able to do better after taking your course than they did before taking it?)

All 16 course supervisors alluded to course content when answering this question. The severity of that allusion varied, however. For example, one calculus supervisor stated that old tests were available for students to serve as content guidelines, thereby affirming the constancy of course design. A supervisor of a mathematics for liberal arts course indicated an overarching goal: students should "appreciate the flavor" of the course content. Another mathematics for liberal arts supervisor, after briskly replying that his students should "pass the course" and struggling with the question about
objectives, added that his goal was to "get their attention." He also seemed pleased about his students enjoying a particular topic he had added to the course.

Several of the supervisors mentioned both content or processes outcomes that aligned well with various standards common to reform paradigms. In particular, supervisors of mathematics courses for elementary school teachers espoused goals that aligned with NCTM Standards. Statistics course supervisors wanted their students to be better at critical thinking, critiquing and collecting data, writing technical materials, and "mathematical fluency." The latter goal, that students should become mathematically fluent, was shared by one of the developmental course leaders as well, and developmental course leaders at both institutions described a main course objective: to prepare students for upcoming courses.

Despite some allusions to goals and objectives that were reform-oriented, however, most supervisors were concerned with the content highlighted in their syllabi. Aware of the seeming dichotomy of values, three instructors discussed the tension between traditional and reform paradigms in terms of classroom practices. A calculus supervisor admitted to mediating occasional battles between instructors trenched in opposing camps. Another interviewee spoke about chronic disagreements he and others who taught the course. According to that instructor, the course supervisor was very content-oriented, while the instructor wished to spend more time on NCTM Standards-oriented pedagogical techniques. Finally, one supervisor's tenuous relationship with reform-based standards was clear by his comments, as was his belief that standards may be for those who have not reached a particular level of mathematical expertise:
I have spent quite a bit of time at conferences discussing [Standards-based issues] with people, and I still do what I think is right. And, yes, it overlaps the standards a fair amount, because, I think any time when you put a fair amount of highly interested people in a room together to [work out] some standards, a lot of them will be right, appropriate. But, if in my own mathematical experience, I've seen something that is effective, useful, and important, and you have a hard time finding it in the standards, I'm going to do it my way. I think a lot of standards are really appropriate for people who are not extremely knowledgeable and experienced, because they need some guidelines, and it certainly is the case that many people teaching high school math do not have an extensive math background, and they need guidelines, and that's perfectly fine.

That supervisor also added that many mathematics faculty saw themselves as experts in their field, one steeped in a tradition and culture. Additionally, he noted that the culture of mathematics was productive, healthy, and inherently beneficial to the disciplines of mathematics and science. In consideration of those remarks, though, the research at hand attempts to answer the question, "Who benefits from that culture?"

To a degree, all course supervisors recommended or, at least, sanctioned the use of technology in their courses. For instance, in developmental and liberal arts mathematics courses at University W students have a choice of calculator-based or non-calculator-based sections. At least one course supervisor offered his instructors micro-lessons on calculator use to pass on to his students, while others assumed students know how to use the technology or will learn to use it on their own. To help his students via course web site, a University W statistics supervisor posted tutorials for different types of calculators there. Several other supervisors integrated spreadsheets and mathematical or statistical software into their courses.

Some supervisors limited the use of hand-held technology in their courses, however. Although at University E calculator use was integrated into the precalculus
curriculum, it was not integrated into the engineering calculus curriculum. As the supervisor of those calculus classes explained, large class sizes prohibited effective use of the hand-held devices. Notably, none of the classes under his supervision had a large lecture format. Students in those calculus classes could use calculators on assignments, but their use was neither integrated into classes nor permitted on exams. Alternatively, at University W using hand-held calculators was integrated into the calculus curriculum, but those devices with symbolic manipulators were not permitted on exams.

The use of cooperative learning in freshman mathematics courses was generally at the discretion of the instructors, according to course supervisors. Only two supervisors mentioned encouraging instructors to incorporate cooperative learning into their activities. The statistics supervisor at University W used group activities in his large lecture classes with some success, he added, but most supervisors indicated that some instructors use group activities and others do not. Several supervisors remarked that their instructors freely made those decisions, and that they would not tell them how to teach, particularly if an instructor was a professor rather than a graduate student.

Other practices associated with reform, like writing about mathematics and assigning in-class projects, were mentioned infrequently. The calculus supervisor at University W acknowledged encouraging her instructors to incorporate writing, projects, technology, and cooperative learning into their practices. As expected, this was also the case in the mathematics courses for elementary school teachers.

Making connections to science and to other areas of mathematics, a desirable reform objective, was a more common theme in comments by interviewees. Developmental course supervisors at both universities discussed a transition away from
symbol manipulation to concepts, and all of the course supervisors alluded to applications inherent in their curricula. In general, then, comments from supervisors about goals, objectives, outcomes, and standards, indicated movement in the direction of reform, even if that movement is neither consistent nor intensive.

3. How do your instructors learn about these goals, outcomes, standards, and expectations?

During the course of the 16 interviews, it became apparent to the researcher that, for graduate students, an association with course supervisors became part of a process of acculturation into teaching mathematics at the university level. Many new graduate students are new to teaching, and their first informal experience with pedagogy comes from the "sit down" side of the classroom. One the other hand, course supervisors are charged with providing new instructors with some formal exposure to pedagogy. Such training may vary with the course supervisor, however, as the following results imply.

At both institutions, instructor recruits learned that weekly meetings were the most common forums for communication. Every course supervisor held those weekly gatherings early in the semester, but meeting frequencies often tapered off later in the term. All 16 course supervisors gave their instructors outlines of content to be covered during the term, and instructors were generally given decision making power regarding a relatively small percentage of the cumulative points for homework, quizzes, labs, or the like. In every instance tests, quizzes, and other course materials served as templates around which instructors built their classroom practices. In some cases supervisors solicited instructor input, but in others supervisors alone wrote exams. Sometimes, as is the case with precalculus and engineering calculus at University E, old exams were
distributed to instructors to aid the acculturation process. Also, at that institution
developmental course instructors received a "long list of rules" by which to abide. Very
often, communication between supervisors and instructors meant e-mail, and nearly every
supervisor mentioned visiting instructors classes at least once, but, possibly, not at all in
the case of veteran instructors. Additionally, one veteran instructor at University W had
organized regular calculus pedagogy meetings for other instructors.

Calculus supervisors at both institutions viewed common finals as a means of
bringing coherence to course content and as a valid form of assessment. The practice of
giving final exams held, as well, for all freshman mathematics courses. At University E
the precalculus course supervisor stated that what students know at the end of the course
is most important and strongly supported rigorous final exams, as did the engineering
calculus course supervisor there. That calculus course supervisor remarked that
uniformity was an issue, regretting his words aloud, as he wondered about the
incongruity of uniformity with other values commonly associated with the university
learning experience.

Of course, supervisors were delighted to get seasoned instructors, as several of
them related to this researcher. At University E, graduate students were expected to work
up through the sequence of courses as part of their training, as one course supervisor
related, although this researcher could find no written evidence of that expectation.
Clearly, however, acculturated instructors were more desirable to course supervisors.
They required less training and generate fewer problems. As in any society, acculturation
means meeting expectations. In this case it meant meeting expectations of the respective
mathematics departments and course supervisors. Whether such acculturation benefited
students, however, is the question of importance here, and one that was considered more carefully in the subsequent Discussion section.

4. How do you and your instructors determine if these goals, outcomes, standards, or expectations for your course are met?

"We don't," was the most common reaction to this question; six of the 16 supervisors admitted that they didn't know if those goals were being met. Two others, who had more positive reactions, stated that gatekeeper tests were the primary assessment tool, and all supervisors indicated that hour exams were still the main form of assessment in their classes. Quizzes were the second most common form of assessment, sometimes mandated by supervisors and sometimes used at the discretion of instructors. Laboratories and software recitations were common to statistics courses at both institutions and in one of the non-engineering calculus courses as well. Such practices tend to provide instructors with alternative assessment data. Homework was collected and graded, generally, at the discretion of instructors and it always counted much less than exams toward a final grade. Only one instructor mentioned that he required instructors to collect and assess all homework. None of the instructors mentioned homework as a means of checking the progress of their students.

For some courses, assessment involved more intricate analyses than the traditional methods cited above. Developmental course supervisors at both institutions kept records on student achievement, and the data they generated guided decision-making about courses. For example, at University E supervisors changed the format of one developmental course from self-study to instructor led after data analysis indicated low course completion rates by students using the self-study method.
During the interview process, anecdotal evidence about student achievement abounded. Although any such data should be subject to scrutiny, its appearance was in itself evidence that course supervisors listened to students and instructors, and thought about issues related to teaching. Concerns about the breadth and depth of student preparation, abilities, and interests, about gateway tests, and about dealing with institutionalized educational systems were common. Supervisors sometimes felt caught in the middle of what they believed and what the institution or department expected, held back by resources, both personal and institutional. As one calculus supervisor, up for tenure that year, self-criticized, "I'm not doing nearly enough." Her comments epitomized a common theme: assessment was not at a level that gained the confidence of many supervisors at the institutions of interest in this study.

5. Is there anything else regarding goals, outcomes, standards, or expectations about your class that you would like me to know?

Interviewees used the opportunity to answer this question in various ways. Although three had no further comments, all others spoke about personal beliefs or their classroom practices, most, anecdotally. One comment about the lack of articulation between courses came from a supervisor of mathematics for elementary school teachers who claimed that his course and its successor lacked coordination, despite attempts to articulate the two. The coordinator for the subsequent course pointed to the lack of time for deep conceptual understanding in the three Standards-based content areas of geometry, probability, and statistics.

After his interview, a statistics supervisor sent alternative assessment materials he used and recommended, while another statistics supervisor commented favorably on
issues that were raised by this researcher, stating "Maybe I should write these down." The calculus supervisor from University E lamented that, despite the honors calculus program, there was really no good forum for gifted mathematics students at that institution. He believed that his honors calculus class served broadly talented students, but not "serious" mathematics students. In a rather long statement, another calculus for technology supervisor at University E claimed 1) that perhaps ten percent of his students finished the course with real understanding of the content; 2) that the importance of mathematics needed to be emphasized outside the discipline itself for students to appreciate it; and 3) that the push to get students through the curriculum had undermined their educational experience. His comments tended to emphasize disappointment in outcomes, even if those statements were based solely on anecdotal evidence. More positively, the precalculus supervisor at University W spoke about her daily use of cooperative learning, but cautioned that doing it well required work, and that effectively supervising that course was an exhausting experience.

**Meaning and Discussion of the Analysis of Data**

**Significance of the Results**

On average, former SIMMS students did not perform as well as their non-SIMMS counterparts on any of the six college entrance exams of interest here or in their first year mathematics classes. Furthermore, those former SIMMS students attending University E scored significantly lower on SAT and ACT mathematics tests and on ACT Science Reasoning and English Tests. Those former SIMMS students attending University W
scored significantly lower than their non-SIMMS counterparts on SAT and ACT mathematics tests only. Also, at each institution and in both Fall 1998 and in Spring the former non-SIMMS students dominated the SIMMS group, measured by their average grades in Developmental, Before Calculus, and Calculus Course Types. Their domination was significant in Fall 1998 at University E and in Spring 1999 at University W. Had the differences between the two groups been confined to measures of mathematical ability, those differences might well have been attributed to curricular effects. However, differences in non-mathematical measures suggest confounding factors. Because the differences were so pervasive, the researcher questioned whether SIMMS and non-SIMMS students came from the same population in terms of general academic ability or some other related fundamental measure.

Former SIMMS and non-SIMMS students did not significantly differ by Course Type in their average mathematics grades, despite the fact that Before Calculus instructors rated their practices as more "SIMMS-like" by virtue of mean Rscores. However, both SIMMS and non-SIMMS students earned significantly better average grades in their Before Calculus and Calculus courses than did those taking Developmental Classes, despite a significantly lower reform self-rating for Calculus instructors when compared with Before Calculus instructors. These findings suggested some confounding in the association between reform practices and course grades, but also pointed to the pitfalls of "gateway" Developmental courses in which students consistently earned significantly lower grades than in either Before Calculus or Calculus courses.
Despite scientifically rich material that pervaded the SIMMS curriculum, former SIMMS students, proportionately, did not choose SMET majors significantly more than did their non-SIMMS counterparts, although these proportions were about 2% higher at each institution.

Results from instructor surveys indicated moderate reform with instructors favoring classroom practices such as using text problems and real-world problems, using calculators, and having students take notes. Moderately frequently, they supported using worksheets, incorporating problem-solving, and writing about mathematics, and having students give presentations. Less frequently they supported using physical models, facilitating cooperative learning, assigning in-class projects, and using computers in class. Before Calculus instructors rated themselves significantly higher than Calculus instructors in terms of reform, while Developmental instructors fell insignificantly between. Even though University W instructors rated themselves significantly higher on average than University E instructors in terms of reform, the return rate of University W survey forms was not high enough to represent that population of instructors. Therefore, no valid comparison between the two populations was possible.

Data from course supervisor interviews suggested that their goals, objectives, standards, and expectations for students were largely conveyed by course syllabi, which tended to be content-oriented. Often the content on the syllabi reflected the chapters of the assigned text. Such an orientation is part of a culture of academic mathematics that was passed on to new instructors and graduate students. Many supervisors visited the classrooms of their instructors, but usually not more than once per semester, and experienced instructors saw supervisors in their classrooms less often or not at all. Exams
and quizzes were pervasive forms of assessment and were used to unifying content goals, although some freshman courses had laboratories and projects as alternative assessments. Calculator use was relatively high, but there was not necessarily any articulation between courses with respect to technology use. Other practices, such as cooperative learning and homework management, were left to instructors. In general laissez faire management of instructors was the rule, with a single obvious exception, although weekly meetings and e-mail communication were common.

Six of sixteen supervisors were quick to admit that they did not know if goals, expectations, standards, or outcomes for their students were met by their courses. However, Developmental Course supervisors kept accurate records of student achievement. Several supervisors expressed dismay about the lack of assessment components that would better inform them, while one instructor emphatically stated that his course was the best that it could be. Several supervisors admitted to disagreements among instructors about classroom practices. Conversely, affirming the prowess of institutional inertia, one engineering calculus instructor admitted to choosing the most widely used text for his course. Supervisors generally expressed concern about student achievement, but they also alluded to barriers to reform, including time, course content, budget constraints, and student ability levels. From these interviews, then, came a portrait of change in the freshman mathematics classroom at the two state universities in Montana. However, adoption of reform practices was neither homogeneous with respect to place, degree, or philosophical grounding, nor was it as intense as it could have been, even though pockets of reform existed.
Supplemental Analysis

Relationship of the Freshman Data to the Research. As in the Core-Plus studies by Bachelis (1998) and Milgram (1999), students who were veterans of the reform curriculum, the SIMMS curriculum in this case, did not perform as well on college-related measures as their more traditionally exposed counterparts. However, the methodological weaknesses of those Core-Plus studies, including self-reporting and lack of representative sampling, undermined the credibility of any findings in that case. Hence, making generalizations of results or rigorous comparisons between the study at hand and the work of Bachelis and Milgram are nearly impossible.

The study at hand benefits from accurate reporting of student scores, unlike the Core-Plus studies, and it also benefits from the analysis of non-mathematical measures of student performance. It was clear from the study at hand that former SIMMS students did not perform as well as their counterparts on any of the performance measures outlined, mathematical or non-mathematical. Many of those differences were statistically significant. As a consequence of the study, consistently lower non-mathematics measures of former SIMMS students, when compared with those of non-SIMMS counterparts, constituted hard evidence of tracking. That is, the population of former SIMMS students may have been of lower average scholastic ability than the population of non-SIMMS students. This question of tracking by ability before curriculum placement deserves dedicated and complete investigation to accurately determine curricular effects as well as pitfalls for students, as described by Vetter (1994).
Curiously, former SIMMS students' choice of SMET majors did not conform to their generally lower performance on college entrance exams or in freshman mathematics courses. They aspired toward the same types of degrees as their non-SIMMS counterparts. Whether the former SIMMS students completed their SMET degrees in parallel with their non-SIMMS peers is another matter and one that deserves more research.

Relationship of the Instructor Survey to the Research. Results from the analysis of instructor responses mirror those of LaBerge, Zollman, and Sons (1997), who created the instructor survey instrument, despite different types of analyses in the two studies. In their study of thirty faculty at eleven institutions, LaBerge, Zollman, and Sons tallied responses of combined Likert categories for each question, while this researcher averaged the responses in categories to obtain graphic profiles for each Institution by Course Type. Regardless of the study, the results indicated that were more likely to take notes (especially more often), do text problems, use real-world problems and do problem-solving, use calculators, and give student presentations in their classes. Alternately, both studies indicated students were less likely to experience cooperative learning, use physical models, work on projects, or use of computers in class. Hence, responses from the LaBerge, Zollman, and Sons study on classroom practices paralleled those of the study at hand.

While freshman mathematics instructors in Montana universities tended toward more frequent use of calculators and real-world problems during class, they less frequently employed many foundational reform practices built into the SIMMS
curriculum: cooperative learning, in-class projects, writing about mathematics, using computers, and using physical materials (MCTM, 1996-1998). However, at both universities the Before Calculus instructors rated themselves higher than the Developmental instructors and significantly higher than Calculus instructors in terms of their reform classroom practices. The ANOVA factorial design, though, indicated that both SIMMS and non-SIMMS students performed significantly better in Before Calculus and Calculus courses when compared to Developmental courses at both universities and in both semesters of this study. Since there was no significant interaction between Course Type and Curriculum, there was not enough evidence to conclude that former SIMMS students performed any better in Before Calculus courses than in either Developmental or Calculus courses, even though instructors rated Before Calculus courses higher on the reform scale. This finding is at odds with that of Allinger, Lott, & Lundin (1998, p. 16) who suggested, without hypothesis testing or a large sample, that former SIMMS students performed better than the general population of freshman college students in Montana in Before Calculus courses.

**Relationship of the Supervisor Interviews to the Research.** The most striking findings from interviews with mathematics course supervisors at University E and University W involved their penchant for content in describing goals, standards, outcomes, and objectives when juxtaposed against their mistrust of their assessment methods. However, many supervisors did not know how effective their courses were. Reforming their courses meant stepping outside the boundaries of the culture of academic mathematics, and those supervisors described various barriers to doing so. LaBerge,
Zollman, and Sons (1997) described barriers similar to those reported in this study from interviews with mathematics faculty (p. 20), finding moderate evidence of reform practices amidst low levels of awareness of the NCTM Standards. Kull (1996) also claimed moderate reform practice in developmental mathematics programs nationwide. Therefore, while Kull and LaBerge, Zollman, and Sons reported some progress in the direction of reform, as indicated by NCTM Standards (LaBerge, Zollman, & Sons, 1997, p. 20 ; Kull, 1996, p. 145), it is clear that barriers associated with the traditional culture of academic mathematics persist.

Such is the case in Montana at University E and University W. Judging from course syllabi, textbook content tends to guide the curriculum, just as Robitaille and Travers (1992, p. 706) caution. Furthermore, on-going classroom practice only partially conforms to reform recommendations. Predictably, then, any significant reform or "deep structural change" (NCTM, 1989, p. 237), probably does not exist within these institutions, although some lower levels of reform (Hall & Hord, 1987, p. 331) exist at both institutions.

Weaknesses, Uncontrolled Factors, and Incongruities in the Study

Several factors potentially weakened this study, including the procurement of Freshman Data and limitations inherent in the research design. First, Freshman Data procurement from University E and University W constituted an uncontrolled factor. From each registrar two cross-matched data sets produced, respectively, an 81% and a 65% return of expected records. However, neither of these return rates was based on head counts of 1998 Montana high school graduates, but on larger pools of in-state
freshmen. Although in subsequent conversations registrars' representatives accounted for
the lower than expected returns, and although two separate data sets survived cross-
matches of common fields, both data returns were substantially lower than expected.
There was, on the other hand, no evidence to suggest that the resultant samples would
leave SIMMS and non-SIMMS counts disproportional. However, if records were
missing, then samples rather than whole populations would become the collective units of
study, and representation of the respective populations, an issue. More positively, the
numbers of SIMMS student records returned were very close to those estimated
beforehand, leading the researcher to be more confident that the numbers of records
returned for the Freshman Data component spanned the populations of students at
University E and University W.

Second, consider two potential design limitations. Emerging from the measures of
college-related performance of the freshmen in this study was a profile of comparatively
weaker mean performance scores for former SIMMS students when compared with non-
SIMMS students. This lower performance profile was detected in all performance
measures and was statistically significant for many, including non-mathematics measures
at University E. Such a profile suggested that the population of former SIMMS students
may have been less academically able before their transition into the SIMMS curriculum,
and the results herein strongly suggest the following hypothesis: less able students were
tracked into the SIMMS curriculum. To attempt to control for academic prowess was not
part of the design of this study, whereas such a design would require data from each
subjects, predating their high school experiences.
If SIMMS students were to outperform their non-SIMMS counterparts in freshman mathematics courses due to their familiarity with classroom practices, this most likely would have been detected in Before Calculus Courses, according to the superior Rscores of those course instructors. This was not the case in either Fall or Spring semesters regardless of institution. Rather, in all four cases, both SIMMS and non-SIMMS students performed significantly better in Before Calculus and Calculus courses than in Developmental courses, and there was no interaction between Course Type and Curriculum Type. It is possible, then, that other confounding factors hid any effects of classroom practice, or that the intensity of reform in freshman mathematics classes was still too low to produce any measurable effect regardless of high school Curriculum Type. In that case, the methodology in this study may have been too course to detect the effects of reform practice, and finer methods could better explain any effects of reform practices.

Alternatively, these results indicated that developmental mathematics courses acted in terms of grades as a sinkhole for freshman students, and factors other than reform practices in the college classroom more strongly affect performance in freshman mathematics courses. Defining those factors and testing conjectures about them should be the work of a follow-up study, keeping in mind the positive effects of a Standards-based curriculum on achievement and self-esteem documented by Hopkins (1998, p. 102).

Chapter Summary

Results from this study comparing college-related measures of former SIMMS and non-SIMMS students came from the analyses of three data sets from each of the two
state universities in Montana. From the first data set, the researcher determined that non-SIMMS students outperformed SIMMS students in all college entrance examinations of interest, and many of the differences were statistically significant. Similarly, former non-SIMMS students had better mathematics grades at both institutions of interest in their first two semesters of residence. However, that difference was significant only in the Fall at University E and in the Spring at University W. There was no significant interaction between Course Type and Curriculum Type. This result implied a lack of evidence for claiming that Developmental, Before Calculus, or Calculus courses differentially favor former SIMMS students over non-SIMMS students. Finally, this researcher concluded that, although former SIMMS students chose two percent more SMET majors than non-SIMMS students, that difference was not statistically significant.

The second data set consisted of responses to the instructor surveys by freshman mathematics course instructors. From that data set the researcher determined that Before Calculus instructors rated themselves significantly more reformed in their classroom practices than did Calculus instructors at both institutions of interest. Also, the self-ratings of developmental instructors fell insignificantly between those of Before Calculus and Calculus instructors. Additionally, although instructors at University W rated themselves significantly higher than those at University E, the return rate of completed survey forms from University W was not high enough render a representative sample from that institution. Hence the insufficient return rate nullifies the comparison. Finally, a profile generated by averaging responses to the instructor survey indicates that instructors favor some classroom practices but not others. Their responses mirror those of instructors who answered the same questions raised by other researchers.
The third set of data came from responses to interviews of freshman mathematics course supervisors at both universities. When queried about course standards, objectives, goals, and outcomes for their students, supervisors alluded to content laden syllabi but also indicated dissatisfaction with traditional assessments as indicators that goals were being met. Certain courses such as statistics, mathematics for liberal arts, and mathematics for elementary school teachers tended to have more of a reformed flavor, while developmental and engineering calculus courses tended to be traditionally taught. Many supervisors alluded to barriers to reform encountered by the researcher in the literature, although interest in student achievement and in reform practices seemed earnest. During the course of the interviews, the researcher noticed the facility with which institutional inertia associated with the culture of academic mathematics sweeps up and carries new instructors and teaching assistants. In any case, the deep change encouraged by reformers has yet to be realized in the mathematics departments at University E and University W.
CONCLUSION TO THE STUDY

Chapter Introduction

The SIMMS curriculum materials embody the NCTM Standards for grades nine through twelve in a learner-focused (Thompson, 1992, p. 136) approach to mathematics. Each module begins with an Exploration or an Activity that is built around a real-world problem. Embedded in the problem-solving process is the use of calculators or computers, and each lesson emphasizes cooperative learning in the exploration and discussion components. The use of physical materials or models is an integral part of many lessons, as are assessments that involve writing about mathematics. Students are encouraged to give presentations of work in progress and presentations of completed projects (MCTM, 1996-1998). These materials, in contrast to those that are more traditional, encourage a method of learning and teaching that is anything but a "linear subject, mainly concerned with mechanistically teaching facts and skills predominately related to number and generally characterized by paper-and-pencil activity..." (Nickson, 1992, p. 103).

In contrast, Thompson (1992) called the traditionalist approach "content focused" (p. 136), driven by the teacher-centered attempts to present content and explain it to students in order to foster conceptual understanding. Hiebert (1997) described to the United States House of Representatives the typical traditional approach in his analysis of
81 videotapes of teaching in the American classroom as part of the Third International Mathematics and Science Study: A Comprehensive Analysis of Elementary and Secondary Math and Science Education.

The American script is easy to describe. Review homework—from the previous day; Demonstrate the procedure to be used for solving the problems for today; Assign similar problems for students to practice at their seats; Assign more similar problems for homework. This kind of mathematics lesson will sound familiar to most readers because it is the script we have been using to teach mathematics for years. In fact, many wonder how else one would teach mathematics. And that reveals a crucial truth—teaching is a cultural activity. It is how we do things in this country, it is deeply embedded in our beliefs about the subject, about how students learn, about how we should teach. This helps to explain why we have been teaching in the same way for so long, but it also means that it will be difficult to change. It is not surprising that previous reforms have had relatively little impact on changing teaching and why the current reforms have not yet taken deep hold. Distributing written documents and advertising them widely is no match for deep-seated cultural practices, despite how good the documents are or how intensive the advertising campaign.

Evidently, the integration inherent in the SIMMS curriculum materials drives the process of teaching and learning in a drastically different manner from the traditional approach suggested by Nickson, Thompson, and Hiebert.

The Classroom Practices Inventory, used to survey freshman mathematics instructors, captured the essence of those process goals embedded in the SIMMS materials. Of course, the separate questions on that instrument were designed to measure practices recommended by various professional organizations, including the National Council of Teachers of Mathematics, the American Mathematics Association of Two-Year Colleges, the Mathematical Association of America. More generally, The Principles of Good Practice for Undergraduate Education (Chickering & Gamson, 1987)
encouraged active learning, cooperative learning, and diverse ways of learning, and those prescriptions intersected with recommendations that were standards-based.

Substantial differences, then, existed between traditional and SIMMS curricula, and those differences naturally led to questions concerning comparative performance of former SIMMS and non-SIMMS students on college related measures.

The Research Questions Revisited

How did 1998 entering college freshman who experienced at least three years of the SIMMS curriculum compare with those students who were veterans of a more traditional high school curriculum? Did reform-type classroom practices of freshman mathematics instructors differ by Course Type? If so, did one type of course favor former students of either high school mathematics curriculum? Operationalizing the first research question involved comparing former SIMMS and non-SIMMS students at the two major Montana universities on three types of measures: college entrance examinations, freshman mathematics grades, and initial choices of majors. Operationalizing the second and third research questions meant surveying freshman mathematics course instructors to find out which, if any, of their classroom practices were recommended by various professional sets of standards and thus inherently "SIMMS-like." In addition to data collected from surveying course instructors, interviews with course supervisors added depth and perspective to a portrait of classroom practice, informing the researcher about goals, expectations, standards, and outcomes for students in freshman mathematics courses. A summary of results follows, and much of the narrative was condensed in Table 21 and Figure 8 of the Implications section.
At both institutions, the former non-SIMMS students outperformed SIMMS students on all entrance examination measures: SAT verbal, SAT mathematics, ACT mathematics, ACT science reasoning, ACT reading, and ACT English tests. Furthermore, the non-SIMMS group performed significantly higher at University E on all of the above tests except ACT reading, and that group performed significantly higher at University W on SAT mathematics and ACT mathematics tests. The pervasively better performance by the non-SIMMS group, particularly on non-mathematical measures, suggested tracking of student subjects in this study. The phenomenon of tracking undermined inferences regarding the effects of the respective curricula on college performance.

Similarly, the non-SIMMS group outperformed the SIMMS group in all three types of freshman mathematics courses, Developmental, Before Calculus, and Calculus, at both universities over Fall 1998 and Spring 1999 semesters. However, the difference was significant at University E only for Fall 1998 and for University W, only in Spring 1999. There was no statistically significant interaction between Curriculum Type and Course Type in any case, so there is no evidence that any of the three freshman Course Types favored either former SIMMS or non-SIMMS students.

Additionally, all students received significantly higher grades in Before Calculus and Calculus courses than in Developmental Courses. Such evidence affirms the gatekeeper character of those Developmental courses.

Although SIMMS students chose two percent more SMET majors than non-SIMMS students upon entering either institution, their choices were not different enough to be statistically significant. Hence, not enough evidence existed to claim SIMMS
students choose science, mathematics, engineering, or technology majors significantly more often than did non-SIMMS students.

Freshman course instructors were surveyed to measure reform practices in their classrooms as a function of Course Type: Developmental, Before Calculus, or Calculus, to see if any of these favored either former SIMMS or non-SIMMS students. At both institutions, Before Calculus instructors rated themselves significantly more reformed than did Calculus instructors, with Developmental course instructors falling insignificantly between. When considered with the results on freshman grades, there was not enough evidence to suggest that reform practices in the college classroom benefited former SIMMS students any more than they benefited their counterparts.

However, it is possible that the intensity of reform practices in classes of surveyed instructors fell well below that experienced by former SIMMS students. The graphs in Figure 8 of the Broader Implications section below suggest that the ideal SIMMS practice profile is markedly more reformed than the profile of even the most reformed Before Calculus group of Instructors. Also, reform profiles at both universities were very similar, affirming the findings from previous research (LaBerge, Zollman, & Sons, 1997; Kull, 1996).

Clearly, instructors favored certain classroom practices more than others, and some of their favorites were reform practices, but many were not. Favorites include note-taking, using problems from the text and problems that they think model real-world phenomena. They tended to use calculators often, but they moderately favored presentations by students. Conversely, they were reticent about using physical materials,
facilitating group work, assigning in-class projects, or having students write about mathematics and use computers in class.

Care is suggested when making inferences about the intensity of problem-solving as a classroom practice. While more simplistic "exercises" require following steps in an algorithm, more complex "problems" require reflection and, perhaps, taking original steps. This researcher did not make this difference clear to survey participants. Therefore, the gap between ideal SIMMS practice and Instructor practice with respect to problem solving may be wider than it appears.

Finally, interviews with freshman course supervisors confirmed the largely content-focused approach to teaching mathematics in freshman courses at both institutions. Often, that content was outlined in syllabi, indexed by chapter numbers in the course texts, which suggested a text-led curriculum. Also, many supervisors were skeptical that their methods helped students meet course objectives, standards, outcomes, and expectations, most of which were content-oriented. Although supervisors seemed interested in student achievement, they also appeared to be chained to the lore, traditions, and practices of the culture of academic mathematics. The interviews did ferret out pockets of reform, even though many courses remained traditionally taught. In general, barriers to reform, as artifacts of institutional inertia, kept supervisors from easily making changes in classroom practice. Such barriers involved constraints on time and other resources and beliefs consistent with the traditions of the culture.
Answering the Research Questions

Former Non-SIMMS students outperformed SIMMS students on all college-related performance measures, including mathematical and non-mathematical college entrance exams and in Fall 1998 and Spring 1999 mathematics grades. The differences were statistically significant in many but not all cases. As a group, freshman course instructors who taught Before Calculus Classes claimed significantly more reform practices in their classes than did Calculus instructors, with Developmental instructors falling insignificantly between. That pattern of reform did not seem to favor performance of either SIMMS or Non-SIMMS students in terms of their grades. Former SIMMS students chose more, but not significantly more SMET majors than non-SIMMS students when entering both institutions of this study.

Broader Implications

Introduction

Three broader implications, stemming from results of this study and its review of the literature, pointed to a need for changes in program assessment, the need for clarifying variables affecting college-related performance, and the need for better articulation between high school and college mathematics programs. First, pervasively better scores by non-SIMMS students suggested tracking of students of higher ability into Non-SIMMS curricula. Second, the more traditional practices of the culture of academic mathematics were very different from those of standards-based reform that SIMMS materials espouse. Although the influence of classroom practice on college freshman
mathematics grades was not clarified in this study, other researchers (Hawkins, 1998; Hurley, Koehn, & Gantner, 1999) claimed positive effects of reform-based practices. More specifically, the graphs in Figure 8 illustrate the gaps in practice profiles between college instructors and the SIMMS ideal. Third, reformers and traditionalists valued algebraic manipulation in curricular content differently, and that difference affects students (Allinger, Lott, & Lundin, 1998). Each of these three broader implications is considered in more detail in what follows.

Pervasive Performance Differences and Early Tracking.

The pervasively better performance by former non-SIMMS students over their SIMMS peers on every college-related measure of this study could be ignored, and that performance difference is clearly evident in Table 21 below. Of particular importance was the fact that, even on non-mathematical tests, the non-SIMMS students outperformed their SIMMS peers. Wide-ranging better performance by the non-SIMMS students suggested tracking by ability early in high school or even before, as occurs often, according to Vetter (1994). Consequently, because tracking of the subjects in this study was likely, a comparison of SIMMS and more traditional curricula with respect to their effects on college-related performance was compromised.

Several hypothetical situations may have led to early tracking of the students subjects of this study. First, those distinguished students who began a traditional algebra track in eighth grade most likely continued that traditional track in high school, never entering the SIMMS track. Second, it is possible that, because of its "hands on" reputation, the SIMMS materials may have been the curriculum of (someone's) choice for
less able students. Further research is needed to determine who makes such choices and why they are made. Third, traditions associated with the culture of the mathematics programs at universities, and of the institutions in this study in particular, may have played a role in the tracking process. That is, students, parents, counselors, and administrators who knew about the traditional culture in university mathematics departments may have advised higher ability students to take traditional mathematics. This hypothesis is bolstered by the fact that many, if not most, larger high schools in Montana ran both SIMMS and traditional track mathematics programs while the subjects of this study were students in them (Diana Paterson, SIMMS Coordinator, personal communication, March 28, 2001).

Table 21. Comparative Results Between Former SIMMS and Non-SIMMS Students On College-related Performance Measures.

<table>
<thead>
<tr>
<th>College-related Performance Measures*</th>
<th>ACT Math</th>
<th>ACT Eng.</th>
<th>ACT Science</th>
<th>ACT Read</th>
<th>SAT Math</th>
<th>SAT Reason</th>
<th>Fall 1998 Grade</th>
<th>Spring 1999 Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>University E</td>
<td>SL</td>
<td>SL</td>
<td>SL</td>
<td>L</td>
<td>SL</td>
<td>SL</td>
<td>SL</td>
<td>L</td>
</tr>
<tr>
<td>University W</td>
<td>SL</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>SL</td>
<td>L</td>
<td>SL</td>
<td>SL</td>
</tr>
</tbody>
</table>

*"L" denotes that SIMMS students had lower mean scores than non-SIMMS students, and "SL" denotes SIMMS students had significantly lower mean scores than non-SIMMS students.

Cultural Differences Between Reformers and Traditionalists.

This study has not answered the question of whether reform-type classroom practices in college mathematics classes favored SIMMS or non-SIMMS students. However, results of this study and its review of the literature evinced the culture clash
between reformers and traditionalists. Consider the recent statement of Madison (2001),
the Chair of the Mathematical Association of America Task Force on Articulation.

The principal cause of the transition problems in US mathematics education has been described as the lack of an intellectually coherent vision of mathematics among the professionals responsible for mathematics education. The sometimes heated and often public disagreements about the nature of mathematics as well as effective ways to teach it have led to a bewildering variety of curricular and pedagogical approaches. (p. 10)

In Figure 8 below the differences in reform practice frequencies between the model SIMMS classroom and freshman mathematics classrooms at University E and University W are clearly evident. The ideal frequencies for the SIMMS model were derived directly from curriculum materials (MCTM, 1996-1998). Even though statistics from this study show that Before Calculus Instructors were significantly more reformed than Calculus Instructors, that group fell far short of the SIMMS ideal. Furthermore, instructors at University E and University W shared the same classroom practice profile, and that profile matched the descriptions of earlier work of LaBerge, Zollman, and Sons (1997). Such differences suggested cultural gaps between those instructors in reformed and traditional camps, with the traditional model conforming to Hiebert's (1997, p. 8) description of the American classroom. Clearly, much of what was expected of SIMMS students was not what was expected of freshman mathematics students at either Montana university in this study.

Despite the fact that results from this study were inconclusive about the effects of reform practices, Hawkins (1998) previously showed positive effects of standards-based high school curriculum on students' ACT mathematics scores and on their self-concepts.
Additionally, in their meta-analysis of college and university calculus programs, Hurley, Koehn, and Gantner (1999) repeatedly cited the benefits of a reformed curricula and practices over those that are more traditional. In light of those studies, the researcher hypothesized that a lack of intensity of reform in the locations of this study, coupled with other confounding variables, may have caused relatively low registration of reform. In any case, the effects of reform-type practices on college-related measures, such as those noted herein, deserve more study.

Figure 8. SIMMS Ideal vs. Combined Instructor Practices by Course Type.
Reformers, Traditionalists and Mathematical Content.

Deserving more consideration in the comparison of the SIMMS and traditional curricula are content differences, particularly the traditional emphasis on algebraic manipulation. Allinger, Lott, and Lundin (1998, p. 29), after interviewing incoming university freshmen who were former SIMMS students, alluded to the disequilibrium suffered by students when first encountering more traditional algebra-centered mathematics courses. While their SIMMS experience may have prepared students for solving problems, that experience was not algebra-centered. Rather, the SIMMS materials concentrate on theme-based modules that integrate mathematical topics. Those materials certainly contain much algebraic content, but they do not share the traditional content-focused organization alluded to by Thompson (1992, p. 136).

The new Principles and Standards (NCTM, 2000) prescribe algebraic manipulation (p. 301), but only as one of multiple prescriptions for the Algebra Standard (p. 296). In any case, it is very possible that algebraic manipulation has become the lynchpin in the traditional mathematics curriculum, much to the detriment of those whose mathematical education has had another focus. Focusing on the need to change the role of algebraic manipulation in the curriculum, the NCTM produced Algebra in A Technological World (NCTM, 1995), an addendum to the NCTM Curriculum and Evaluation Standards for School Mathematics (NCTM, 1989), from which the following statement comes.

Perhaps more than any other area of school mathematics, the study of algebra is bound to change dramatically with the infusion of currently available and emerging technology. What was once the inviolable domain of paper-and-pencil manipulative algebra is now within easy reach of
school-level computing technology. This technology demands new visions of school algebra that shift the emphasis away from symbolic manipulation toward conceptual understanding, symbol sense, and mathematical modeling (p. 1).

Developmental Courses at both institutions of this study were algebra courses, and students earned significantly lower grades in them than did those students in Before Calculus and Calculus courses. Although it is clear from supervisor interviews that reform has hacked its way into these institutions' algebra domains, its presence was not commanding. As an indicator, calculators were excluded from exams at University E, and those calculators that perform algebraic manipulation were banned from exams at University W. While the use of technology represents only one indicator of reform, it is a pivotal one, since that technology can accomplish algebraic manipulation fast and accurately. It is clear, however, that in the minds of some instructors, that computational power also undermines a part of algebraic-manipulation that they think needs to be mastered by students.

Adelman's academic intensity variable (Adelman, 1999, Introduction, p. 2) was a powerful enough predictor of college graduation to deserve Standards-based redefinition. Consider his conclusion from his findings.

When academic intensity and quality of one's high school curriculum is such a dominant determinant of degree completion, and both test scores and (especially) high school grade point average or class rank are so much weaker contributors to attainment, college admissions formulas that emphasize test scores and (especially) high school grade point average or class rank are likely to result in lower degree completion rates (p. 4).

That variable, however, based on the prevalent traditional mathematics paradigm during the 1980-1993 period of his study, rewarded traditional courses like Algebra I and
Algebra II with extra "quality" points (Adelman, 1999, p. 3). Reformed courses could not have been so rewarded, even if they showed up in Adelman's data. Therefore, the definition of Adelman's intensity variable aligned well with traditional courses, but it could not accurately reflect intensity for mathematics courses that are more modern. Indeed, the relatively high predictive value of the traditional version of Adelman's intensity variable further validates the strong association among traditional high school mathematics courses, traditional college mathematics courses, and college achievement. Conversely, a modern version of Adelman's intensity variable would provide focus for mathematics educators and lead the way toward reconciliation between traditional and reform camps.

What These Implications Mean.

This study showed that former SIMMS students entering University E and University W did not perform as well as their non-SIMMS counterparts on college entrance exams or in their freshman college mathematics courses, although both groups initially chose about the same types of majors. This study did not conclusively demonstrate why this happened. Instead, the results herein combined with the literature search made a strong argument relating four variables to that performance, all within the context of change and cultural gaps. Those variables are: 1) tracking during or before high school; 2) inherent practices within the culture of mathematics; 3) the role of algebra within high school and college mathematical curricula as related to the culture of mathematics; and 4) the intensity of the mathematical curriculum.
No model for predicting mathematics performance in college will be accurate if it does not account for those variables. Furthermore, unless some reconciliation among mathematicians and mathematics educators takes place with respect to content, practice, and assessment goals, a comprehensive model for predicting achievement is impossible; indicators for achievement must follow from consensus. Evidently, as this study shows, consensus regarding goals, outcomes, standards, and expectations is not a matter of fact in Montana or, for that matter, in other states.

The science of change theory predicts educational evolution in stages, and the state of mathematics reform, judging from results herein, is not at a stage that satisfies many mathematics educators. More than research, then, is needed to end a fray where all players, especially students, are caught in the middle. More positively, this researcher has suggestions, gleaned from careful analysis of the data that might make the current state of affairs better. Therefore, from the results of this study and from research cited herein, come the following recommendations.

**Recommendations: Research and Reconciliation.**

1. Research on tracking must happen at the high school level and before. This research is important for at least two reasons. First, without data that follows students through their school years, no conclusions can be made about equity or quality of programs. Second, without data from students, common predictive achievement indicators, such as scores on college entrance exams,
cannot account for differences in high school programs, and those differences evidently exist. Minimally, high schools should have assessment programs that provide administrators with information about their mathematics programs and their students' achievement, even after they graduate.

2. Changing the way mathematics is taught means changing the cultures of mathematics education and academic mathematics. Values, beliefs, and practices must change. Substantial conflicts now exist, arising from ideological differences about how to present mathematics, what to present, and how to assess what is learned. Much research already exists on best practice, not that new research should be unwelcome. What is needed more, though, is a community effort among mathematicians and mathematics educators to work through key issues. As with any cultural differences, awareness, communication, and dialogue are suggested as vehicles by which resolution of differences may take place. More to the point, any solid program that mixes high school and college mathematicians and mathematics educators in a substantial, truly meaningful fashion will help reconcile differences. Better articulation of programs will be a byproduct of the understanding that follows. National, state, and local programs that encourage articulation are not just recommended; they are critically needed now. An indicator of reconciliation will surely come from comments of college freshmen about the continuity of their mathematical experiences, those comments eventually replacing the disparaging remarks about the lack of coherence and articulation in current mathematics programs.
3. This study has not taken on the whole issue of mathematics curricular content and achievement. However, it was clear from the results of course supervisor interviews and from the literature that, at the very least, the role of algebra in high school and college mathematics is changing. The algebraic content, as emphasized by the NCTM Standards, was not the algebra that was emphasized in college classes.

   Millions of students are entering college, veterans, perhaps, of NCTM Standards-based programs that emphasized decreased attention to certain algebraic topics (NCTM, 1989, p. 187). How the colleges handle this decreased attention to once-valued topics is unclear and begs for research. Meanwhile, the current split in ideologies begs for mathematics programs that mediate the goals and objectives of reformers and traditionalists.

   Also clear from this study is the fact that performance in developmental algebra courses is significantly below that of other Course Types at the two Montana Institutions of this study. Any future study that purports to examine that phenomenon must take into account the role of course placement, usually dependent on placement tests or entrance exam scores. In any case, research in placement and achievement in Developmental courses is strongly recommended for both institutions. A major goal of that research should be to provide recommendations for high school mathematics educators in Montana, so that prospective students might be better prepared for college mathematics.

4. A modern definition of academic intensity of the high school mathematics curriculum has two important advantages. First, any good definition teases out
important goals, and that process focuses on important mathematics. Redefinition on academic intensity should help bring consensus to a fragmented mathematics community. Second, if the definition is valid, it describes an indicator or variable for prediction of college performance, and a powerful predictor, as Adelman showed (1999, Introduction, p. 2). Because defining what is meant by academic intensity is both a value judgement and culmination of research, redefinition via research on academic intensity is recommended.

5. A note on methodology associated with the Freshman Data component of this study is important here. This study was largely descriptive, a fact-finding mission. If the object of this study were to construct a linear model for the prediction of freshman grades, more sophisticated statistical tests would have been in order. That next level of research would have employed the power of a general linear model to ascertain which, if any, of the dependent variables were more meaningful in their prediction of freshman performance. To be reliable, however, such a model would need to account for the tracking of students, requiring accurate measures academic ability of students early in high school or before. Future studies might build such a model.

6. Finally, assessment should guide practice at all levels. This limited study did not do justice to that assertion. The value of studying assessment at the classroom level and at the program level in future research of this type cannot be overemphasized. Indeed, a full study of classroom practice and how it relates to student assessment and program assessment in light of reform is recommended. Such a study might even initiate a
move toward coherence in a portrait of college classroom teaching currently fragmented by disparate traditional and reform goals.

Chapter Summary

This chapter began with a summary of how the SIMMS curriculum differed from more traditional curricula. The research questions were revisited and answered. Major findings showed that former non-SIMMS students outperformed SIMMS students on mathematics and non-mathematics measures. Before Calculus and Calculus students received significantly better grades than Developmental students. The degree of reform in freshman mathematics courses favored neither SIMMS nor non-SIMMS students, and neither group chose significantly more SMET majors.

Three important broader implications of the results led to six specific recommendations for research and reconciliation. Evidence for tracking begged for research and assessment at the high school level. Cultural differences among traditional and reform camps demanded reconciliation efforts consistent with any cultural gap to unify goals and practices. The changing role of algebra in the mathematics curriculum demanded research into course placement and agreement on content, especially into developmental courses. The next level of research needed to consider better predictive models of college-related performance. Academic intensity in the spirit of Adelman needed redefinition in terms of the NCTM standards to force out common goals for a new curriculum and to create a powerful variable for the prediction of college achievement. Lastly, in light of
reform, research on assessment at the classroom and program levels was called for to bring coherence to a portrait of mathematics teaching, fragmented by the disparate goals and practices of reformers and traditionalists.
REFERENCES


American Association for Higher Education in Washington, DC, June 8-12, 1996. (ERIC Document Reproduction Service No. ED 397 769)


Milgram, R. J. (1999). Outcomes analysis for Core Plus students at Andover: One year later. Retrieved February 21, 2000 from the world wide web:


APPENDICES
APPENDIX A
The Classroom Practices Inventory

Directions: For the following 12 questions, please circle the response that best indicates how frequently YOUR STUDENTS participate in the indicated activities **IN CLASS**.

**How frequently do your students...**

<table>
<thead>
<tr>
<th>Question</th>
<th>Never</th>
<th>Seldom</th>
<th>Sometimes</th>
<th>Most of the time</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ...work exercises or problems from the textbook?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. ...use physical materials or models?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. ...work exercises or problems from a worksheet or handout?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. ...learn mathematics from real-life applications, concepts, or procedures?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. ...work in groups?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. ...make conjectures and explore problem-solving methods?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. ...use calculators?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. ...work on mathematics projects or open-ended investigations?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. ...take notes while the teacher lectures?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. ...write about mathematics?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. ...use computers?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. ...present or discuss solutions to mathematics problems?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*(LaBerge, Zollman, & Sons, 1997)*
Freshman Course Supervisor
Interview Protocol

The main idea of these interviews is to elicit responses from course supervisors about the goals, outcomes, expectations, or standards they have for the students in the courses they supervise. These will be compared to those recommendations from the literature.

Introduction

My research considers the performance of incoming college freshmen. Part of that research measures their performance on college admissions tests and in their college mathematics classes. It is also concerned with how their high school mathematics courses affect their performance on these measures and how they affect their choices of majors. Your comments concerning the goals, outcomes, standards, and expectations of your students establish a framework within which their performance may be examined. Your statements will remain confidential, and no full transcript will be published. If I use quotations, they will be anonymous.

Questions

1. Do you have a written set of goals, outcomes, standards or expectations against which students' achievement may be measured in \(<\text{math XXX}>\)? (Are these included in the course syllabus or some other document?) If so, may I have a copy?

2. What goals, outcomes, standards, and expectations for the students in \(<\text{math XXX}>\) are stressed by you and your instructors? (In other words, what do you want your students to be able to do better after taking your course than they did before taking it?)

3. How do your instructors learn about these goals, outcomes, standards, and expectations?

4. How do you and your instructors determine if these goals, outcomes, standards, and expectations are met in \(<\text{math XXX}>\)?

5. Is there anything else regarding the goals, outcomes, standards, or expectations about \(<\text{math XXX}>\) that you would like me to know?

Conclusion

Thank you for your time and responses. Please feel free to contact me if you have any questions about this interview or the content thereof.
APPENDIX B
The NSF Drivers of Reform

D1. Implementation of comprehensive, standards-based curricula as represented in instructional practice, including student assessment, in every classroom, laboratory, and other learning experience provided through the system and its partners.

D2. Development of a coherent, consistent set of policies that support provision of high quality mathematics and science education for each student; excellent preparation, continuing education, and support for each mathematics and science teacher (including all elementary teachers); and other learning experiences provided through the system and its partners.

D3. Convergence of the usage of all resources that are designed for or that reasonably could be used to support science and mathematics education--fiscal, intellectual, material, curricular, and extracurricular--into a focused and unitary program to constantly upgrade, renew, and improve the educational program in mathematics and science for all students.

D4. Broad-based support from parents, policy makers, institutions of higher education, business and industry, foundations, and other segments of the community for the goals and collective value of the program, based on rich presentations of the ideas behind the program, the evidence gathered about its successes and its failures, and critical discussions of its efforts.

D5. Accumulation of a broad and deep array of evidence that the program is enhancing student achievement, through a set of indices that might include achievement test scores, higher level courses passed, college admission rates, college majors, Advanced Placement Tests, taken, portfolio assessment, and ratings from summer employers, and that demonstrate that students are generally achieving at a significantly higher level in science and mathematics.

D6. Improvement in the achievement of all students, including those historically underserved.

(MCTM, 1997)
Equity Principle: Mathematics instructional programs should promote the learning of mathematics by all students.

Curriculum Principle: Mathematics instructional programs should emphasize important and meaningful mathematics through curricula that are coherent and comprehensive.

Teaching Principle: Mathematics instructional programs depend on competent and caring teachers who teach all students to understand and use mathematics.

Learning Principle: Mathematics instructional programs should enable all students to understand and use mathematics.

Assessment Principle: Mathematics instructional programs should include assessment to monitor, enhance, and evaluate the mathematics learning of all students and to inform teaching.

Technology Principle: Mathematics instructional programs should use technology to help all students understand mathematics and should prepare them to use mathematics in an increasingly technological world.

Five standards describe the mathematical content that students should learn:

- Number and Operation
- Patterns, Functions, and Algebra
- Geometry and Spatial Sense
- Measurement
- Data Analysis, Statistics, and Probability

Five standards describe the mathematical processes through which students should acquire and use their mathematical knowledge:

- Problem Solving
- Reasoning and Proof
- Communication
- Connections
- Representation

(NCTM, 1998)
SMET and Non-SMET Majors at University E and University W

<table>
<thead>
<tr>
<th>SMET</th>
<th>Non-SMET</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>University E</strong></td>
<td><strong>Non-SMET</strong></td>
</tr>
<tr>
<td>Abused Land Rehabilitation</td>
<td>Ag Business-Agribusiness Mgmt</td>
</tr>
<tr>
<td>Ag Operations Tech</td>
<td>Ag Business-Farm &amp; Ranch Mgmt</td>
</tr>
<tr>
<td>Animal Sci-Livestock Mgmt</td>
<td>Ag Ed-Broadfield Teaching</td>
</tr>
<tr>
<td>Animal Sci-Science</td>
<td>Ag Ed-Extension</td>
</tr>
<tr>
<td>Bio Sciences-Biology</td>
<td>Ag Ed-Teaching</td>
</tr>
<tr>
<td>Bio Sciences-Biomedical Sci</td>
<td>Art-Art Ed K-12 Broadfield</td>
</tr>
<tr>
<td>Bio Sciences-Fish &amp; Wildlife</td>
<td>Art-Studio Arts</td>
</tr>
<tr>
<td>Bio Sciences-Teaching</td>
<td>Business</td>
</tr>
<tr>
<td>Biotechnology-Animal Systems</td>
<td>Business-Accounting</td>
</tr>
<tr>
<td>Biotechnology-General</td>
<td>Business-Finance</td>
</tr>
<tr>
<td>Biotechnology-Plant Systems</td>
<td>Business-Management</td>
</tr>
<tr>
<td>Chemical Engr</td>
<td>Business-Marketing</td>
</tr>
<tr>
<td>Chemistry-Biochemistry</td>
<td>Economics-General Economics</td>
</tr>
<tr>
<td>Civil &amp; Ag Engr-Bio-Resources</td>
<td>Elementary Ed-Early Childhd Ed</td>
</tr>
<tr>
<td>Civil &amp; Ag Engr-Civil Engr</td>
<td>Elementary Ed-K-10</td>
</tr>
<tr>
<td>Computer Engineering</td>
<td>Elementary Ed-Mathematics</td>
</tr>
<tr>
<td>Computer Science</td>
<td>Elementary Ed-Special Ed</td>
</tr>
<tr>
<td>Construction Engineering Tech</td>
<td>English</td>
</tr>
<tr>
<td>Earth Sciences</td>
<td>English-Literature</td>
</tr>
<tr>
<td>Earth Sciences-Geology</td>
<td>English-Teaching</td>
</tr>
<tr>
<td>Electric &amp; Electron Engr Tech</td>
<td>General Agriculture</td>
</tr>
<tr>
<td>Electrical Engr-Electric Engr</td>
<td>HHD-Child Dev &amp; Family Sci</td>
</tr>
<tr>
<td>Environmental Design</td>
<td>HHD-Community Health</td>
</tr>
<tr>
<td>General Engineering</td>
<td>HHD-Family Science</td>
</tr>
<tr>
<td>HHD-Exercise Science</td>
<td>HHD-Home Econ Educ/Ext</td>
</tr>
<tr>
<td>HHD-Food &amp; Nutrition</td>
<td>History-History</td>
</tr>
<tr>
<td>HHD-Health Enhancement</td>
<td>History-Teaching</td>
</tr>
<tr>
<td>HHD-Pre-Physical Therapy</td>
<td>Media &amp; Thtr Arts-Mo/Vid/Thr</td>
</tr>
<tr>
<td>Horticulture-Horticulture</td>
<td>Media &amp; Thtr Arts-Photography</td>
</tr>
<tr>
<td>Horticulture-Landscape Design</td>
<td>Modern Languages-Commerce</td>
</tr>
<tr>
<td>Horticulture-Turfgrass Science</td>
<td>Modern Languages-Spanish</td>
</tr>
<tr>
<td>Industrial &amp; Management Engr</td>
<td>Music Ed-Schl Music K-12 Brd</td>
</tr>
<tr>
<td>Mathematics-Mathematics</td>
<td>Music Ed-Studio Teaching</td>
</tr>
<tr>
<td>Mathematics-Teaching</td>
<td>Political Science</td>
</tr>
<tr>
<td>Mechanical Engineering</td>
<td>Psychology</td>
</tr>
<tr>
<td>Mechanical Engineering Tech</td>
<td>Secondary Ed-Social Study Bd</td>
</tr>
<tr>
<td>Academic Program</td>
<td>Academic Program</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Microbiology-Microbiology</td>
<td>Sociology-Anthropology</td>
</tr>
<tr>
<td>Physics-Physics Professional</td>
<td>Sociology-Justice Studies</td>
</tr>
<tr>
<td>Pre-Architecture</td>
<td>Sociology-Sociology</td>
</tr>
<tr>
<td>Pre-Nursing</td>
<td></td>
</tr>
<tr>
<td>Pre-veterinary Medicine</td>
<td></td>
</tr>
<tr>
<td>Psychology-Applied Psych Opt</td>
<td></td>
</tr>
<tr>
<td>Range Science</td>
<td></td>
</tr>
<tr>
<td>Secondary Ed-Gen Sci Brdfd</td>
<td></td>
</tr>
<tr>
<td>Soils-Soil &amp; Enviro Science</td>
<td></td>
</tr>
<tr>
<td>Technology Ed-Ind Tech Opt</td>
<td></td>
</tr>
<tr>
<td><strong>University W</strong></td>
<td></td>
</tr>
<tr>
<td>Biology</td>
<td>Anthropology</td>
</tr>
<tr>
<td>Chemistry</td>
<td>Art</td>
</tr>
<tr>
<td>Computer Science</td>
<td>Business Admin.</td>
</tr>
<tr>
<td>Forestry</td>
<td>Drama</td>
</tr>
<tr>
<td>Geology</td>
<td>Economics</td>
</tr>
<tr>
<td>Mathematics</td>
<td>English</td>
</tr>
<tr>
<td>Medical Technology</td>
<td>Fine Arts</td>
</tr>
<tr>
<td>Microbiology</td>
<td>French</td>
</tr>
<tr>
<td>Pre-Engineering</td>
<td>Health, Human P.</td>
</tr>
<tr>
<td>Physics</td>
<td>History</td>
</tr>
<tr>
<td>Pre-Medical</td>
<td>Japanese</td>
</tr>
<tr>
<td>Pre-Nursing</td>
<td>Library Science</td>
</tr>
<tr>
<td>Pre-Pharmacy</td>
<td>Music</td>
</tr>
<tr>
<td>Pre-Physical Therapy</td>
<td>Pre-Communications</td>
</tr>
<tr>
<td>Resource Conservation</td>
<td>Performance</td>
</tr>
<tr>
<td>Wildlife Biology</td>
<td>Philosophy</td>
</tr>
<tr>
<td></td>
<td>Pre-Journalism</td>
</tr>
<tr>
<td></td>
<td>Pre-Law</td>
</tr>
<tr>
<td></td>
<td>Pre-Psychology</td>
</tr>
<tr>
<td></td>
<td>Pre-Education</td>
</tr>
<tr>
<td></td>
<td>Political Science</td>
</tr>
<tr>
<td></td>
<td>Range Resource Mgmt.</td>
</tr>
<tr>
<td></td>
<td>Sociology</td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
</tr>
<tr>
<td></td>
<td>Social Work</td>
</tr>
</tbody>
</table>