EFFECTS OF MAYFIELD'S FOUR QUESTIONS (M4Q) ON NURSING STUDENTS' SELF-EFFICACY AND MULTIPLE-CHOICE TEST SCORES

A Dissertation

Presented to

The Faculty of the School of Education

Liberty University

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Education

by

Linda Riggs Mayfield

July 2010

Effects of Mayfield's Four Questions (M4Q) on Nursing Students' Self-Efficacy and Multiple-Choice Test Scores

by Linda Riggs Mayfield

APPROVED:

COMMITTEE CHAIR	Karen L. Parker, Ed.D.
COMMITTEE MEMBERS	Carol Ann Moseley, Ph.D.
	Jeffrey Savage, Ed.D.
CHAIR, GRADUATE STUDIES	Scott B. Watson, Ph.D.

Abstract

Linda Riggs Mayfield. EFFECTS OF MAYFIELD'S FOUR QUESTIONS (M4Q) ON NURSING STUDENTS' SELF-EFFICACY AND MULTIPLE-CHOICE TEST SCORES. (Under the direction of Dr. Karen Parker) School of Education, June, 2010.

This study examined the effects of being taught the Mayfield's Four Questions multiple-choice test-taking strategy on the perceived self-efficacy and multiple-choice test scores of nursing students in a two-year associate degree program. Experimental and control groups were chosen by stratified random sampling. Subjects completed the 10-statement General Self-Efficacy survey and two additional statements based on guidelines by the survey authors and self-efficacy theorist Albert Bandura. The intervention consisted of weekly half-hour sessions to learn and practice M4Q. Survey data was statistically analyzed by group and by statement. Multiple-choice test scores were analyzed by group, level, test and test group. Statistically significant differences were found between the final experimental (n = 16) and control (n=17) groups in post-intervention perceived self-efficacy, change in perceived self-efficacy, specific self-efficacy factors, scores on the final standardized nursing test and the mean score of the predictor tests for the licensure examination. A significant correlation was found between the number of M4Q teaching sessions attended and 2nd year students' grades on two teacher-prepared tests. Results indicate the M4Q can have a positive effect on both self-efficacy and multiple-choice test performance. Suggestions for further research were discussed.

Dedication

"Not unto us, O Lord, not unto us, but to Your name give glory,

because of Your mercy, because of Your truth"

Psalm 115:1 (NKJV)

Earning the Doctor of Education degree at this stage of my life would not have been possible without the loving, unwavering, unconditional, prayerful support of my godly husband of more than 40 years, Brian K. Mayfield, and constant encouragement from him and our children, Melissa J. Mayfield Ledbetter and her husband, Nate; Joshua K. Mayfield and his wife, Brooke; Brian Caleb Mayfield, and Sarah E. Mayfield. The unfailing love of them and our grandchildren Selah Grace, Shiloh Sky and Levi Mayfield Ledbetter; and Noah Kent Mayfield has been a constant inspiration.

Acknowledgements

I am increasingly grateful for my parents, now with the Lord, Maynard E. and Estel M. Riggs, for encouraging me to love and serve God, to seek the wisdom and knowledge that leads to Truth, and to be a life-long learner. They would have been delighted to see this day. The dissertation process would not have been possible without the enthusiastic willingness of my husband, Brian, to make my deadlines and priorities *our* priorities. Our children, Melissa, Joshua, Caleb and Sarah have been constant sources of love and affirmation. Prayer support, interest, and kind words from many friends, especially Holly Rich, Joyce Farlow and my pastor, Mickey Farlow, have supported me in the process again and again.

My committee chair, Dr. Karen Parker, has been an invaluable resource, encourager, "level head," colleague and friend, for which I am most grateful. My advisor, Dr. Scott Watson, facilitated the program transition and maximized my opportunities. In his role in the dissertation process, he forced me to think in different ways. I deeply appreciate Dr. Carol Ann Moseley, Professor Emeritus at Blessing-Rieman College of Nursing, who served on my committee and gave generously and patiently of her time and statistics expertise. I am also grateful for Dr. Jeffrey Savage's wide-ranging knowledge, attention to detail, and competent service on my committee.

I am extremely grateful to Blessing-Rieman College of Nursing and its corporate entity, Blessing Corporate Services, for institutional policies and provisions that allowed me to complete my degree. I am grateful to Pamela Brown, RN, PhD, President, and Ann O'Sullivan, RN, MSN, CNE, Associate Dean, for their unwavering support. I am thankful for the faculty, staff and students at Blessing-Rieman who have worked with me

and encouraged me, particularly the members of the Research Committee and the librarians, as well as my student workers, Amanda Lutener and Binoy Thomas, RN, BSN. I am also deeply grateful for program director Anne Riggs, RN, MSN, and the faculty, staff and nursing students at Hannibal-LaGrange College for their kindness, generosity and support in hosting and facilitating this research.

Contents

Title……. i

Approvals...……. ii

Abstract…….…….……..….. iii

Dedication…….……..……... iv

Acknowledgements.…….…….. v

Contents…… vii

Tables and Figures..... xvi

1. Introduction…….……... 1

Need for the Study…….……. 2

Healthcare Significance…….……. 2

Nurse Shortage Significance…….….…… 2

Demographics…….…….……... 2

Causes…….…….……... 3

Ageing and Attrition……... 4

Enrollment and Image…….…… 4

Faculty Shortage…….…… 4

Activism…….……..… 5

Missing Factor…….…… 5

High-Stakes Tests in Nursing Education……..…….. 6

Limits to Usage…….…….. 7

Pass Rates…… 8

Institutional Risks…… 8

Test-Taking Efficacy…….…….. 9

One Model for Success: Mayfield's Four Questions.……..…… 9

The Problem…….....…… 9

New Premise.……..…… 10

New Strategy……...… 10

Testing the Strategy…….…...…. 11

Potential……...…..... 12

Focus of the Study…….…..…..... 12

Problem Statement……... 12

Research Questions…….....…. 13

Null Hypotheses…….....….. 13

Operational Definitions of Hypotheses Terms……..… 13

Theoretical Support……...…... 14

Social Learning, Agency and Self-Efficacy Theories…….. 14

Instrumentation……...….…… 14

Operational Definitions of Literature Terms…...……..….…… 16

Significance of the Study……....…… 19

Self-Efficacy Literature…….……..……... 19

Test-Taking Literature…….....….. 20

Summary of the Need for the Study……..……...…… 21

2. Review of the Literature…….……. 21

Bandura's Social Cognitive, Agency and Self-Efficacy Theories……..…..…... 21

Social Cognitive Theory……..……... 21

Development of Social Learning Theory……..…..… 22

Sears to Bandura…….…..… 22

Dyadic to Triadic……..... 24

Agency Theory…….……..….. 25

Core Properties of Agency …….. 25

Intentionality..... 25

Forethought..... 26

Self-Reactiveness..... 26

Self-Reflectiveness..... 26

Personal Agency ……..……...…..... 26

Spiritual Agency…….….. 27

Spiritual Agency in Testing…….. 27

Bandura's Fortuitous Determinants……..…….29

Biblical Christian View of Agency……..…...….. 31

Will..... 31

Predestination..... 32

Consequences..... 33

Self-Efficacy Theory……...…..... 34

The Model…….....…..… 34

Domains……...…....…..... 34

Sources of Self-Efficacy…….……...….. 35

Influences of and Responses to Bandura's Theories……...……... 36

Model Development……..……..…. 36

Quantitative Research……...…….. 36

Self-Efficacy Assessments……..…….... 37

The General Self-Efficacy Scale……..……... 38

Self-Efficacy and the College Student……...….. 39

Self-Efficacy and Student Affairs…….....…. 40

Critics of Bandura's Theories……..... 40

Stress, Anxiety, Coping and the Test-taker……..... 40

Definitions…….....…..... 41

Stress in Testing……..... 42

Trait and State Anxiety…….....… 43

Coping by the Test-Taker……..…... 44

Sense of Coherence (SOC)……..…....…. 45

Self-Efficacy and Coping…….……...… 45

High-Stakes Tests……....… 46

No Child Left Behind……...……..... 46

Challenges…….…...…… 47

Limitations in the Literature……...….…. 48

Second-Language Students……..... 48

All Self-Reported Data....…….…..... 48

Focus on Psychological Stress……..... 49

High-Stakes Tests in Nursing Education……...…….. 49

Test Banks…….....… 50

Instructor-Written Questions……..…..… 51

HESI……...…. 52

Use…….....… 52

High-Stakes……...….... 53

Concerns……...……..... 54

Conclusions……..……..... 55

Institutional Responses to Nursing Students' Testing Needs……... 55

Best Practices……...…. 55

High-Stakes Tests and Remediation……..... 56

Conclusions from the Literature on High-Stakes Tests……..... 57

Strategic Test-Taking……...…. 58

Test-Taking Support……..... 58

Commonalities……..… 59

Lack of Evidence..... 62

Differences……..…..…. 62

Reading or Not Reading the Options..... 63

Purposeful Advice..... 63

"Unique" Advice..... 63

Nursing Compared to Non-Nursing……..……..... 64

Age of Advice……...… 65

Academic Folk Wisdom…….…….…. 65

A Potential Remedy…….….…… 66

Conclusion and Recommendation…….….... 67

3. Methodology…….……... 68

Research Design…….….. 68

Sampling Procedures…….…... 69

Site……..…….…. 69

Population…….…… 69

Sampling…….…….. 70

Subject Protection…….…… 71

Pre-Emptive Protection…….…… 72

Institutional Review Boards…….….…… 72

Data Collection and Protection…….…… 73

Instrumentation…….…… 75

Intervention…….….…… 76

Content…….……. 77

Conceptual Design……..…….. 78

Threats to Internal Validity……..…. 81

Subject Mortality..... 82

History..... 82

Instrumentation..... 82

Data Analysis…….....…….. 83

Comparison of Proposed and Actual Data Analysis Design…….….. 83

Descriptive Treatment……. 85

Survey Responses..... 85

Test Scores..... 86

Omnibus Run and Multiple Group Analysis...…… 86

Factor Analysis…… 86

Comparisons and Correlations Among the Variables…..... 86

Structural Modeling……. 86

Summary of Data Analysis Design.....…… 88

4. Results……. 89

Descriptive Statistics…… 90

Null Hypothesis #1: Self-Efficacy Scale Results..... 90

Visual Inspection Descriptive Results…….....……. 92

Statistical Analysis: *T*-Tests Results..... 92

Changes in Experimental Group's Scores..... 92

Changes in Control Group's Scores..... 93

Comparisons Between Groups..... 94

Experimental Group Results…….....……. 97

Null Hypothesis #2: Multiple-Choice Nursing Test Grades..... 97

Comparisons Between the Groups on Formative Tests..... 97

Comparisons Between the Groups on Summative Tests..... 98

Differences in HESI Scores..... 98

5. Summary and Discussion…….…….....100

Introduction…….....100

Statements of the Problem…….....100

Review of the Methodology…….....102

Discussion of the Conceptual Design of the Intervention.....103

Summary and Discussion of Statistical Results…….....105

Conclusions Regarding Null Hypothesis #1: Self-Efficacy.....105

Outcomes.....105

Summary of Survey Data Supporting Rejection

of the Null Hypothesis.....105

Discussion..... 106

Conclusions Regarding Null Hypothesis #2:

Multiple-Choice Nursing Test Performance Results..... 106

Summary.....106

Discussion.....106

Outcomes..... 107

Summary of Survey Data Supporting Rejection

of the Null Hypothesis..... 107

Discussion..... 107

Interpretation of the Findings..... 110

Relationship of the Current Study to Previous Research.....110

Limitations.....112

Size.....112

Institutional Climate.....114

Recommendations for Further Research.....114

References.....116

Appendices.....129

A. IRB Approval 786.010810: Effect of the Mayfield¹s Four

Questions (M4Q) Test-taking Strategy on Nursing Students'

Perceived Self-efficacy and Multiple Choice Test Scores.....129

B. Blessing-Rieman College of Nursing IRB Approval.....130

C. Informed Consent Form..... 131

D. Pre-and Post-Intervention General Self-Efficacy (GSE) Survey.....135

E. Handout for Mayfield's Four Questions..... 137

Tables and Figures

Table 1. Multiple-Choice Questions in Nursing Test Banks (Masters et al, 2001)..... 50

Table 2. Test-taking Advice in Selected Texts1995-2010:

Factors in Common with M4Q..... 61

Table 3. Test-taking Advice in Selected Texts1995-2010: Contrasts to M4Q..... 62

Figure 1. Conceptual Design: Mayfield's Four Questions (M4Q)..... 81

Table 4. Multiple-Choice Tests Data Sources..... 90

Table 5. Demographics for Experimental and Control Groups..... 93

Table 6. Experimental Group Pre- and Post- Self-Efficacy Survey Question Means... 98

Table 7. Control Group Pre- and Post- Self-Efficacy Survey Question Means..... 101

Table 8. ANOVA for Between-Groups Comparison

of Pre- and Post-Intervention Surveys..... 104

Table 9. Correlations Between Level 2 Subjects' Number of Sessions

and HESI Scores..... 109

Table 10. ANOVA Pearson Correlations Between Experimental Group

Number of M4Q Sessions and Grades on Instructor-Prepared

Multiple-ChoiceTests..... 117

CHAPTER 1: INTRODUCTION

This dissertation describes a research study that examined the effects of being taught a specific multiple-choice test-taking strategy on the self-efficacy perceptions and multiple-choice test scores of nursing students. The time frame of the study's intervention was the spring, 2010, academic semester, and all the subjects were enrolled in an associate degree nursing program.

Chapter 1 introduces the study; explains the social, institutional, and individual contexts that support the need for the study; and presents an overview of the study's theoretical foundation. Chapter 2 is an explication of the findings in the literature relative to four broad constructs: (1) Bandura's Social Learning Theory, Agency Theory, and Self-Efficacy Theory as foundations for the study, (2) psychological responses to learning stressors such as high-stakes tests, (3) the role of high-stakes tests in nursing education, and (4) strategic test-taking in education in general and nursing education in particular. Chapter 3 describes the methodology of the experimental design for this research. The GSE instrument used for the assessment of self-efficacy is adequate in design, validity and reliability to contribute to evidence-based literature. Chapter 4 is the discussion of the findings that resulted from the statistical analysis of the data from both the self-efficacy surveys and the multiple-choice test scores. A dual-method approach allowed the potential for benefit to be examined from more than one perspective. Chapter 5 is the interpretation and application of the findings of the study.

Need for the Study

Healthcare Significance

Americans consider healthcare to be one of the most important social issues facing the United States. In at least three national polls conducted between August, 2009 and April, 2010, healthcare ranked first in non-economic issues of concern to Americans, and ranked second only behind the economy in all issues (Saad, 2009; Bloomberg National Poll, 2010; Fox News Poll/Opinion Dynamics, 2010).

Healthcare is a complex issue that includes not only the questions of how much it will cost and who will pay for it, but also who will provide it. Registered nurses comprise the largest group of healthcare professionals in the United States (Buerhaus, Staiger, & Auerback, 2000). The number of nursing positions in the U.S. is expected to increase at a rate of 22% in a ten-year period from 2,618,700 in 2008 to 3,200,200 in 2018. These figures represent a growth rate that is faster than the average for all occupations (Registered Nurses, 2009). About 60% of the 2.6 million jobs for registered nurses are in hospitals, with most others employed in physicians' offices, home health care services, nursing care facilities and employment services (Registered Nurses, 2009). There are not enough nurses available to fill the openings, however, and addressing the need for more nurses is a multi-faceted challenge.

Nurse Shortage Significance

Demographics. The current shortage of professional registered nurses in the United States is unprecedented and is expected to continue (Unruh & Fottler, 2005). A shortage exists in every major venue of healthcare in which nurses are employed. The American Hospital Association reported 116,000 unfilled registered nurse (RN) positions

in the nation's hospitals at the end of 2006 (Nursing Shortage Fact Sheet, 2009). In 2007, the average vacancy rate for all RN positions was 8.1% (The 2007 State of America's Hospitals—Taking the Pulse, 2007). By mid-2008, the American Health Care Association reported 19,400 vacancies in RN positions in long-term care facilities nationwide (Nursing Shortage Fact Sheet, 2009). The overall nursing shortage in one state, Iowa, for example, is currently 8%, and is projected to increase to 27% within the next ten years (Iowa: Need for Nurses, 2010).

On a separate web page on the corporate website, health products giant Johnson & Johnson eloquently summarized the need for more nurses:

> We are acutely aware that today our nation faces the most profound shortage of nursing professionals in its history. A shortfall of nearly half a million registered nurses as well as a serious shortage of nursing educators over the next decade threatens to disrupt both the quality and availability of health care services. Enrollment in nursing degree programs has fallen 17% over the last five years. Adding to the problem, registered nurses are leaving the profession at an ever-increasing rate. All of this is happening as the largest single generation in our nation's history reaches its age of greatest health care need (DiscoverNursing.com, 2010).

Causes. Several factors summarized by Johnsons & Johnson have converged to create the nursing shortage, some sociological and some demographic. Goodin's 2003 meta-analysis of literature addressing the shortage identified four broad categories of causes: the ageing of the nursing workforce, attrition of practicing nurses due to work climate, declining enrollment in nursing programs, and the public image of nursing. A

closely-related cause is the shortage of nursing faculty.

Ageing and attrition. The people at the leading edge of the Baby Boomers, a large demographic group of people born from 1946 to 1964, are reaching retirement age. With increased age often comes increased need for medical/nursing attention, which requires more nurses to provide the care. Many nurses and nursing instructors who are working today are themselves Baby Boomers whose retirement will leave voids in the both the practice and academic workforce. Nurses are also leaving the profession due to widely-observed and well-documented workplace issues (Nursing Shortage Fact Sheet, 2010; Wood, 2010).

Enrollment and image. Before the social revolutions of the 1960s, few professional opportunities for women existed other than teaching and nursing. Today, women can enter any profession, and that abundance of opportunity drains potential nurses into other professions. Nursing is not viewed as being as socially desirable as some other fields that are now open to women, so the pool of individuals who want to become nurses is smaller, even while the need is greater (Goodin, 2003).

Faculty shortage. The shortage of nursing faculty is acute for some of the same reasons the shortage of nurses exists, particularly ageing and attrition. Some college nursing programs have been forced to limit their enrollments due to their inability to acquire more instructors. Others have restructured their curricula to allow the same number of instructors to teach more students. Another tactic is hiring qualified nurses to teach classes or clinicals in their specialty as part-time adjunct faculty. While helpful as stopgap measures now, none of these models is adequate to meet the shortage, which is expected to increase (Registered Nurses, 2009; Nursing Shortage Fact Sheet, 2010;

Gardyasz, 2010).

Activism. The need for more nurses is so great, very diverse entities in American society have undertaken unique and innovative initiatives to address it. One example took place in Iowa; the American Association of Retired Persons partnered with a coalition called Iowa Needs Nurses Now (2010; Gardyasz, 2010). Iowa Needs Nurses Now invited volunteers to participate in an effort to increase opportunities for potential nurses to both matriculate and complete degrees in nursing. Another example was corporate and national: Johnson & Johnson launched a high-profile television advertising campaign that dramatically promoted nursing and nursing instruction as careers to be valued and considered (DiscoverNursing.com, 2010).

Missing factor. The literature on the nursing shortage focused on choosing nursing as a career, enrolling in a nursing program, and perseverance in practice as a professional nurse after completing a program. The important issue of the academic success of the enrolled nursing student between matriculation and graduation was minimally addressed. Nursing programs are traditionally structured and somewhat lock stepped: only after one level is mastered is a student permitted to progress to the next, and there are several levels. A significant percentage of the students who begin an associate's or baccalaureate degree program in nursing are not successful in mastering every level, completing the program and graduating. Research is needed that examines what happens during enrollment.

High-Stakes Tests in Nursing Education

One significant hindrance to completing a nursing degree program is that students must pass a series of formative standardized multiple-choice tests, each one often a

"gatekeeper" to being permitted to continue in a class, a level, or the degree program. The most widely-used of the high-stakes standardized tests for discrete nursing content are published by Elsevier, and are referred to as HESI examinations. HESIs are discussed in Chapter 2.After completing a degree, graduates must pass the National Council of State Boards of Nursing examination, commonly referred to as the NCLEX (NCLEX statistics from NCSBN, 2010), in order to be licensed. Without a license, the graduate nurse is not a registered nurse, and cannot practice as a nursing professional. Each of these formative and summative gatekeeper tests in nursing education is high-stakes. The American Educational Research Association (AERA) (2000) defined high stakes as those that "carry serious consequences for students or for educators" (para. 3).

The NCLEX is an interactive computer examination administered in proprietary test sites, such as Sylvan Learning Centers, by individual appointment. Each examinee's first question is randomly-selected by the computer program from a medium level of difficulty. The choice of each subsequent question is based on whether the response to the previous one was correct or incorrect, resulting in a vast array of possible question sequences. The test is psychometrically designed so an examinee who does extremely well or extremely poorly will pass or fail after completing 75 questions. Students who alternate between answering correctly and incorrectly can take a maximum of 265 questions before passing or failing.

Limits to usage. The American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) jointly published *Standards for Educational and Psychological Testing*, to guide educators in the appropriate use of high-stakes assessments (AERA

Position Statement on High-Stakes Testing in Pre-K-12 Education, 2000). One of their guiding criteria echoed by other educators is that high-stakes decisions should not be based on one test (Carpenter, 2001). That caveat is not widely followed (Spurlock, 2006) and it is particularly not followed in nursing education: gatekeeper tests are typical at all levels. The rationale for including so many high-stakes assessments is that nursing education prepares students for a profession in which their family members, neighbors, and strangers will entrust their lives to them. Their education must be thorough, and the standards of accountability must be exemplary.

Cautions have been put forth related to such emphasis on high-stakes assessments, however. Since the NCLEX is so high-stakes, the usually-disreputable practice of teaching to the test is mandatory, in a sense. The NCLEX is designed to determine the minimum level of safe practice, so students and their instructors must constantly speculate what knowledge the writers of the NCLEX questions deemed to be necessary for safe practice when they designed the test bank and what content they omitted because they considered it helpful to know, but not necessary. The sequence of questions presented to the examinee is randomly selected by the computer program, so one student's "weak spot" in knowledge can show up again and again in that individual's test questions, while another student might see only one or two questions on the topic. Having a stronger knowledge base in every other construct may not help the final score enough to take it to the passing level. NCLEX examinees must be knowledgeable on a very broad range of nursing content topics, and have the test-taking skills to represent their knowledge accurately.

Pass rates. Data posted by the National Council of State Boards of Nursing

(NCSBN) show that a significant number of examinees are not successful on their first attempt at passing their NCLEX examinations. In the first quarter of 2010, 10% of the 35,259 U.S.-educated first-time examinees were not successful. In 2009, 11.6% of the 134,709 U.S.-educated examinees were not successful. In 2008, 13.3% of the 129,121 U.S.-educated first-time examinees were not successful. From January, 2008 through March, 2010 a total of 33,151 potential nurses who had graduated from a degree program were not successful on their first attempt at passing the NCLEX. Many remediated and made second and third attempts, but fewer than 56% of the U.S.-educated graduates who took the NCLEX a second or third time in that time span passed it (NCLEX Statistics from NCSBN, 2010).

Institutional risks. Using the NCLEX as the ultimate high-stakes test for the student who has already successfully completed a degree in nursing creates a painful winnowing process. Doing so maintains the high standards necessary for the profession, but further reduces the ultimate number of college nursing graduates who become professional nurses. Nursing programs which cannot maintain a sufficiently-high NCLEX first-time pass rate by their graduates face penalties ranging from an official warning to withdrawal of program accreditation. Not only for themselves, but for their institutions, students must be maximally-prepared to be academically successful in their nursing programs and pass the licensing examination on the first attempt. A strategy enabling more nursing students to be successful on multiple-choice tests would address several of the contributors to the nursing shortage. Unsuccessful test-taking by students and unsuccessful NCLEX attempts by graduates would diminish, and fewer colleges would be at risk of losing their accreditation because of low first-time pass rates.

Test-taking efficacy. Students who persist in the challenges of a nursing degree program must believe they are capable of passing the requisite tests; that is, believe in the adequacy of their own knowledge base and in their own self-efficacy as test-takers. They should reasonably expect to achieve the passing scores if they have the necessary content knowledge and strategic learning skills to do so. The data indicates, however, that many do not.

One Model for Success

The problem. At a small, single-purpose college of nursing affiliated with a major medical center in the Midwest, students who were not successful on formative tests have been referred to the Associate Faculty for Academic Enhancement for help for many years. Every year numerous students stated that they had no difficulty narrowing the multiple-choice test questions' options down to two, but then they always seemed to choose the wrong one. The admission requirements for the college were sufficiently selective that every student enrolled should have been able to learn the content. The curriculum was thorough and innovative, and the college's library had been recognized as the best medical research library in the state. The faculty members were highly qualified, both by education and experience. Evidently, some other factors were at play in these students' inability to pass the standardized multiple-choice tests.

Students often could read a test question silently or orally, but not be able to accurately restate what was being asked. They often could not identify and explain what curriculum content the question was assessing—a skill, a theory, an application of knowledge, an understanding of a process, or something else. An observable pattern was particularly manifested by the fourth-year, senior-level students. Based on their own

advanced knowledge and clinical experience, they often read into questions and added layers of more complex or complicated meaning than was stated and/or intended in the syntax of the questions.

New premise. The researcher in this study and author of the dissertation was the Associate Faculty for Academic Enhancement (AE) of the college—the educational learning specialist on a faculty otherwise comprised entirely of nurses. The search began in the AE Office for the reason well-educated students were having so much difficulty passing nursing tests. The inquiry included searches in the literature that are discussed in Chapter 2 of this paper. Qualitative evidence accumulated from years of records kept in professional practice. A successful new approach to test-taking emerged and was ultimately formalized as Mayfield's Four Questions (M4Q).

New strategy. The M4Q test-taking strategy employed a sequential critical thinking process about each test question. Students were instructed not to look at the options offered as possible answers. Four questions about the item were considered that progressively directed the thought process toward the correct answer. Student were challenged to determine (1) what the question was really asking, (2) what the question was really trying to determine is known, (3) what level of learning was required to answer correctly, and (4) what they think the correct response will be. Then the options were scanned, not for consideration, but only to locate the one that had been identified in the fourth step. (See Appendix C).

In tests and measurements nomenclature, incorrect response options are *distractors*. Consistent with the terminology, test questions often appeared to have been written to intentionally distract students from the correct answer. When students were

not attentive to the incorrect answers, however, the distractors lost their power to distract. A strategy that facilitates attending to what one knows to be true (correct response), rather than to several options that are not true (distractors), could have positive results for students and for those who administer assessments. Students who did know the content being assessed would answer more questions correctly, and the discrimination effectiveness of the multiple-choice instruments would be heightened, as well. M4Q is thoroughly delineated and modeled in Chapter 3: Methodology.

Testing the strategy. After reviewing a few years of mounting anecdotal evidence, the Academic Committee of the college mandated that every student who experienced academic difficulty be required to access remediation that included M4Q. Students who had failed a course due to not achieving the minimum percentage correct on multiple-choice tests came for weekly sessions in which they practiced the M4Q on questions from published test banks. Anecdotal data indicated that every student's test scores improved, and in a few cases the improvement was so dramatic that the students went from earning failing grades to not just passing, but earning very high grades. As evidence of the efficacy of the strategy accumulated, the academic progress of students who did and did not access the help was compared. For the spring semester of the year preceding this research, every student on probation who made appointments and learned the M4Q strategy was successful, and every student on probation who did not make appointments and learn the strategy was unsuccessful.

Conducting experimental research that excluded a control group from learning the strategy would not be ethical at the college in which many students had already improved their test scores by using M4Q. The next logical step was to formally, experimentally

test the strategy at another college. The director of the nursing program at a small, Christian liberal arts college affiliated with the Southern Baptist Convention was approached about hosting a research project in her department in the Spring 2010 semester. She enthusiastically agreed and facilitated the logistics of the research project.

Potential. The results of an experimental study testing M4Q have the potential to be a valuable addition to the literature and practice of nursing education. An increase in the number of successful nursing students could positively impact the nursing shortage and the quality of health care in the U.S. Further research is also warranted that will test the M4Q strategy with other student populations that take high-stakes multiple-choice tests.

Focus of the Study

Problem Statement

Nursing students are required to be successful on numerous high-stakes multiple-choice tests in order to complete an academic program, graduate, and become registered nurses. At a time in which there is an unprecedented nursing shortage, great numbers of potential nurses are lost due to inability to pass the tests.

Research Questions

The Research Questions addressed by this study were:

(1) Do nursing students who are taught the Mayfield's Four Questions (M4Q) multiple-choice test-taking strategy perceive higher self-efficacy than nursing students who are not taught the strategy?

(2) Do nursing students who are taught the M4Q multiple-choice test-taking strategy achieve higher multiple-choice test scores than nursing students who are not

taught the strategy?

Null Hypotheses

The null hypotheses of this study were:

(1) Nursing students who were taught the Mayfield's Four Questions (M4Q) test-taking strategy will not perceive a post-intervention level of self-efficacy that is higher than the level perceived by nursing students who were not taught the strategy.

(2) Nursing students who were taught Mayfield's Four Questions (M4Q) test-taking strategy will not earn multiple-choice test scores that are higher than the multiple-choice test scores of nursing students who were not taught the strategy.

Operational Definitions of Hypothesis Terms

- *Mayfield's Four Questions (M4Q):* the test-taking strategy being validated in this study.
- *Nursing students:* first- and second-year students enrolled in the associate degree nursing program at a particular Christian liberal arts college in the Midwest.
- *Intervention:* teaching Experimental Group students the M4Q and meeting with them to practice it using published test questions.
- *Were taught*: met with the researcher for at least two appointments for the purpose of learning Mayfield's Four Questions (M4Q).
- *Were not taught*: met with the researcher for fewer than two appointments for the purpose of learning M4Q, either by choice or by random selection for the Control Group.
- *Multiple-choice test scores*: the percentage grades the Control Group and

Experimental Group participants earned on all formative and summative multiple-choice tests that were administered in their nursing courses during the Spring 2010 semester.

- *Perceived self-efficacy*: the Control and Experimental groups' average scores on a pre- and post-survey consisting of the General Self-Efficacy Scale (Jerusalem & Schwarzer, 1994) with two additional questions.

Theoretical Support

Social Learning, Agency and Self-Efficacy Theories

A study of a test-taking intervention requires theoretical support that elucidates beliefs about the student's pro-active role in learning. Three of Albert Bandura's theories provide that theoretical support. Bandura's Social Learning Theory defines the individual's role as an active learner. The Agency Theory aspect of Social Learning posits that the individual is an active force in his or her own learning. The Self-Efficacy Theory aspect of Agency Theory explicates the how the individual's ability to address and cope with stresses results in a belief that one can succeed, and that the belief is a factor that supports the success. The three related theories are particularly congruent with the study of a test-taking strategy and are thoroughly discussed in Chapter 2.

Instrumentation

The instrument chosen for a pre- and post-intervention survey was the General Self-Efficacy Scale (Jerusalem & Schwarzer, 1994). Scherbaum, Cohen-Charash, & Kern (2006) found that "over the past 20 years, self-efficacy has become one of the most widely studied variables in the educational, psychological, and organizational sciences" (p. 1047). Two colleagues at the Free University of Berlin, Matthias Jerusalem and Ralf

Schwarzer, developed the General Self-Efficacy Scale, a brief, broad assessment instrument. It was published in German about 1980: sources variously list the date as 1979 and 1981 (RCMAR Measurement Tools, 2006) and translated into English by the original authors in 1993 (Schwarzer & Jerusalem). It has subsequently been subjected to validation studies in at least 25 countries in a wide range of constructs and is available online for researchers' use in 27 languages (RCMAR Measurement Tools, 2006; Jerusalem & Schwarzer, 2008). The Resource Centers of Minority Again Research evaluation of the instrument stated that its strengths are ease of administration and interpretation. No weaknesses were noted.

Scherbaum, Cohen-Charash and Kern (2006) completed a complex statistical comparative study of three General Self-Efficacy assessment instruments: the General Self-Efficacy Scale, the General Perceived Self-Efficacy Scale, and the New General Self-Efficacy Scale. The researchers determined that "for all three measures, the responses demonstrated satisfactory internal consistency […] and the measures were statistically significantly and positively correlated" (p. 1053). The limitations they identified—no test for order effects, the potential for fatigue effects and the sample being constituted of students--were all addressed. They concluded that "the items on these measures function fairly well and demonstrate many desirable psychometric properties" (p. 1058).

Bandura (2006), Pajares (1996), and others caution that self-efficacy is most accurately assessed within a context or specific situation. Bandura formally addressed the issue of measuring self-efficacy via instrumentation in his "Guide for Constructing Self-Efficacy Scales" (2006). In a caveat consistent with that viewpoint, Jerusalem and

Schwarzer state that "in most applications it is necessary to add a few items to cover the particular content of the survey or intervention" (2008, p. 1). Based on these caveats, two questions were added to the GSE which allowed it to become specific to the study. The adapted GSE was considered adequate and appropriate for this study.

A search of a group of psychology databases (PsycArticles, PsycBooks, PsycExtra, PsycInfo 1887-Current, and ProQuest) using *General Self-Efficacy Scale* as the keywords produced 30,135 results in September, 2009. In May 2010, the search was replicated and produced 30,370 results. Studies were reported in North America, Asia and the Middle East. An increase of 235 published articles in eight months indicates ongoing worldwide research about and/or using the GSE instrument.

Operational Definitions of Literature Terms

A consideration of important terms in the study, stated in the sequence in which concepts are presented, demonstrates the logical flow of themes addressed in the review of the literature.

- *Stress*: the physical or psychological perception that the demands being placed upon the individual are beyond the individual's ability to react without a negative response. It is interactive between the individual and the stressor. Lazarus and Folkman (1984) called it "a particular relationship between the person and the environment that is appraised by the person as taxing or exceeding his or her resources and endangering his or her well-being" (p. 19).
- *Stressful situations:* occasions in which the individual perceives the reality as beyond normal coping ability, in other words, that the situation will "tax or exceed the person's available resources" (Bolger, 1990, p. 525). Test-taking is

often a stressful situation.

- *Coping:* summoning cognitive and other resources to manage stress and the causes of stress (Lazarus & Folkman, 1984). Applying test-taking strategies one has learned is an example of coping.
- *Agency:* intentional pro-activeness. Agency is "the capability of individual human beings to make choices and to act on these choices in ways that make a difference in their lives" (Martin, 2004, p. 135). Bandura, who developed agency theory, said that "to be an agent is to influence intentionally one's functioning and life circumstances" (2006, p. 164) or "the course of environmental events" (2008, p. 86).
- *Strategies:* cognitive or behavioral reactions intended to positively address a given stressor within its context (Zeidner, 1995). Strategies are intentional, pro-active and agentic, that is, they require purpose and initiative by the individual.
- *Coping strategies***:** ways the individual chooses to deal with stresses. There are three types of coping strategies: (1) problem-focused, (2) emotion-focused, and (3) avoidance (Zeidner, 1995, p. 290). Problem-focused coping strategies direct attention to the issue at hand and facilitate cognitive decisions about it. Studying for a test is a problem-focused strategy. Emotion-focused coping strategies are psychological responses, such as denial or blaming. Avoidance as a coping strategy is diverting attention from the stressor in order to not deal with it. Avoidance strategies include procrastination, rationalization and substitution of activity.
- *Coping resources:* an individual's repertoire of strategies to deal with

various stressors. Coping resources can be inherent, as in a personality characteristic, or learned, such as stress-management techniques. They can make it possible for "individuals to handle stressors more effectively, experience fewer or less intense symptoms upon exposure to a stressor, or recover faster from exposure" (Zeidner, 1995, p. 280).

- *Sense of Coherence (SOC):* an individual's perception of his or her own ability to marshal diverse appropriate resources to deal with stress (Cohen, Ben-Zur & Rosenfeld, 2008). SOC is an awareness that can include confidence if the perception is that one's coping resources are adequate for the given stressor.
- *General Self-Efficacy:* the belief that one can be effective. After realizing one's ability to gather resources (SOC), one must believe he or she can use the resources to competently cope with the general life stressors being faced. General self-efficacy specifically refers to overall life effectiveness, not efficacy in specific or targeted areas (Luszczynska, Scholz & Schwarzer, 2005).
- *Specific Self-Efficacy:* also called contextual or situational self-efficacy, one's belief in the ability to be effective when focused on a specific task (Lippke, Wiedemann, Ziegelmann, Reuter & Schwarzer, 2009; Luszczynska, Scholz & Schwarzer, 2005). Self-efficacy as a test-taker is an example of Specific Self-Efficacy.
- *High stakes test:* an assessment whose consequences have the potential to effect a significant life change or direct consequences for an individual, program, and/or institution (Spurlock, 2006). The summative examination in nursing is high-stakes for all three: the student, the nursing program, and the college.

Significance of the Study

Self-Efficacy Literature

A significant and growing body of literature addresses the perceived self-efficacy of individuals. Emory University posted a website entitled "Student Research on Self-Efficacy: A Community of Scholars" that identified hundreds of student-designed self-efficacy studies in progress or completed by September, 2005. Although 16 studies were identified in the category *Academic Self-Efficacy*, none addressed test-taking. While 38 studies addressed *Medical Issues and Settings, Clinical, Health*, none focused on nursing students. This study made a contribution in both areas.

A search of a group of psychology databases (PsycArticles, PsycBooks, PsycExtra, PsycInfo 1887-Current, and ProQuest) using *General Self-Efficacy Scale* as the keywords produced 30,370 results in May, 2010. A great preponderance of the published articles addressed self-efficacy that was measured by one instrument in particular: Jerusalem and Schwarzer's General Self-Efficacy Scale (GSE) (2008). The research software identified clusters based on themes. Clustered results indicated that many of the studies dealt with children, psychological issues, and the validity and contructs of the test itself. Not enough of the articles were identified as specifically education-related for the software to create a clustered results group about education. Following further refinement of the search, only two articles were found that specifically addressed educational test-taking, and both of them were studies of test anxiety. None were found that examined the relationship between self-efficacy and the acquisition of a strategic test-testing skill. This study also made a contribution to that need in the literature.

Test-Taking Literature

A significant body of literature addressed the psychological aspects of taking tests, but it almost exclusively focused on the stress and anxiety associated with test-taking (Bolger, 1990; Blankstein, Flett, & Watson, 1992; Zeidner, 2007; Thompson & Gaudreau, 2008) and the coping activities that were used or recommended to deal with stress (Bolger, 1990; Davis, DiStefano, & Schutz, 2008). A significant majority of the strategies in print that did address strategic test-taking, both in generic and nursing-specific texts, advised students to study the distractors when deciding upon a response. That was the opposite approach from the fourth step in M4Q. The published material on test-taking strategies was, with few exceptions, not supported by published research. This study made a contribution to the need for evidence-based practices in the literature.

Summary of the Need for the Study

Nursing students' ability to pass high-stakes tests is a requisite skill at every formative level of a nursing degree program, and is required again after graduation in order to pass the NCLEX, the summative and ultimate high stakes test in nursing. A test-taking strategy that enabled nursing students to perceive themselves as possessing efficacy in test-taking and taught them to be more successful on their numerous high-stakes tests would benefit the students, their colleges, and ultimately the nation's healthcare system. It could have a far-reaching practical social application in addressing the current acute nursing shortage. The M4Q test-taking strategy also has particular significance because it challenges the majority of the advice about the best way for students to approach multiple-choice test questions. This study was needed to support the literature of academic self-efficacy with evidence-based research on strategic test-taking.

CHAPTER 2: REVIEW OF THE LITERATURE

The research conducted for this dissertation focused on several components that are not necessarily otherwise related, but are integrally related in this study. Broad but discrete concepts were addressed in the literature review in consideration of that reality, and their relationships to each other are delineated. The broad concepts are: (1) Bandura's Social Cognitive, Agency and Self-Efficacy Theories; their development, assessment tools, and application to test-taking; (2) the psychological constructs of stress and coping, specifically in the context of test-taking; (3) high-stakes tests and the specific challenges of nursing students as high-stakes multiple-choice test-takers, and (4) the strategies recommended in the literature as coping mechanisms to promote effectiveness in multiple-choice test-taking.

Bandura's Social Cognitive, Agency and Self-Efficacy Theories

The theoretical framework for this study was Albert Bandura's Social Cognitive Theory. The Agency aspect of Social Cognitive Theory was isolated as relevant. The Self-Efficacy Theory component of agency was explored and applied to the academic activity of test-taking.

Social Cognitive Theory

Albert Bandura, David Starr Jordan Professor Emeritus at Stanford University, immigrated to the United States from Canada to attend graduate school. He earned an M.A. in 1951 and a Ph.D. in 1952 in clinical psychology from the University of Iowa. Following a one-year internship at the Wichita Guidance Center, Bandura accepted a faculty position at Stanford University in 1952, where he has taught, conducted research and theorized to the present (Bandura, 2006). He is widely recognized as the creator of

Social Cognitive Theory, considered one of the most important and influential theories in psychology and education (Pajares, 2004).

Development of Social Learning Theory. Beginning in 1935, researchers at the Yale Institute of Human Relations addressed the goal of establishing a legitimate and unified science of human behavior. They became known as the Yale Group. Their work bridged the gaps between the detached observation of Freudian analysis, the narrowness of the stimulus-response mechanisms of behaviorists, and the planned intervention that is inherent in education.

The Yale Group's first published research was a collaboration of five authors, including Robert Sears. In 1941, two of the five--Neal Miller and John Dollard--published the first major presentation that used the terms *social learning theory,* a book entitled *Social Learning and Imitation*. The authors posited that the primary learning model for individuals is *social modeling,* or imitation of observed behaviors and their resulting outcomes as they are performed by others. Their concept of social learning was supported by research in young children, and demonstrated that learning activities can be implemented by planned intervention. Planned intervention is foundational for educational research (Grusec, 1999; Pajares, 2004) in general and this study in particular.

Sears to Bandura. Albert Bandura was exposed to the ideas of the Yale Group when he was a student at the University of Iowa and when he subsequently became an instructor at Stanford University, where Robert Sears became chair of the psychology department. At Sanford, Bandura and graduate student Richard Walters collaboratively merged the earlier ideas of psychoanalysis with new social modeling ideas about learning, and used interview as the means of gathering data, rather than observation, to

conduct their research. By the time they published *Social Learning and Personality Development* in 1963, they had moved beyond the psychoanalysis foundation altogether, and believed they had developed a new concept they named the *sociobehavioristic approach* (Grusec, 1999).

The interventions based on Bandura's emerging theories of learning were so successful that in 1969, he published *Principles of Behavior Modification*, a book he said "addressed the influential role of cognitive, vicarious, and self-regulatory mechanisms in human adaptation and personal and social change" (Bandura, 2006, p. 62). The researcher was a participant in the research process, and not only an observer. Bandura's next areas of research carried the concepts of social learning from models even further, adding stronger elements of self-awareness, self-processes, self-motivation, and in particular, self-efficacy.

In 1977, Bandura published *Social Learning Theory*, in which he stated that "theorists who exclude the capacity for self-direction from their view of human potentialities restrict their resources to external sources of influence," and added that, "people are not simply reactors to external influences" (Preface, p. vi). Bandura had begun his studies with the theoretical framework of Social Learning Theory, the term coined by the members of the Yale Group, but by the mid-1980s, he recognized that numerous other kinds of theories—drive theory, expectancy theory, operant conditioning theory, etc., were all being included under the umbrella of social learning theory in the growing research and resulting literature. His own learning theory was broader in application than the specific foci of the elements social learning had come to include, so

he elected to differentiate his theory by naming it Social *Cognitive* Theory (Bandura, 2006).

Dyadic to triadic. Bandura's Social Cognitive Theory added to the dyadic approach of Sears, granting that there is influential activity both ways between an individual and the environment, but Bandura found the use of the dualistic, dyadic approach to be divisive and insupportable (Bandura, 1977, 2002). Bandura rejected a limited behavioristic foundation in which the learning is accomplished only in response to a stimulus. By adding one more element—the autonomous cognitive actions of the individual-- to Sears's dyad, Bandura developed a triadic approach (Pajares, 2003) and backed it with extensive experimental testing (Bandura, 2007).

Bandura's studies demonstrated that the individual develops environment perception and self-perception based on cognitive representations that include (1) response-outcome expectancy (projecting anticipated results of chosen actions), (2) standards for evaluating one's own reactions (developing a sense of values), and (3) self-efficacy perception (believing in one's own ability to accomplish or achieve). Learning results from a constant interaction between the three elements—the individual, the environment and the behavior, not only the environment and the behavior (Bandura, 1977). He summarized the rationale for the triadic approach by explaining that his theory "neither casts people into the role of powerless objects controlled by invironmental forces nor free agents who can become whatever they choose. Both people and their environments are reciprocal determinants of each other (Bandura, 1977, p. vii). The triadic viewpoint is pivotal to the belief that one can be taught to take tests strategically.

Agency Theory

Social Cognitive Theory was formalized as the unexpected result of a study that examined the relationship between psychological and medical variables related to phobias (Bandura, 2007). Every subject in the study was able to overcome the phobia he or she possessed, so the intervention was extraordinarily successful, but in follow-up interviews, Bandura discovered a broader outcome. Subjects who had successfully learned to regulate a phobic negative response were willing and able to subsequently try addressing other stressors that had previously elicited negative responses and were being avoided (Pajares, 2004). They thought (cognition) they decided to act (agency), and were able to do it (efficacy). That finding provided the foundation for Bandura's subsequent development of Agency Theory.

Core properties of agency. Bandura's theory identifies four core properties of agency: (1) intentionality, (2) forethought, (3) self-reactiveness, and (4) self-reflectiveness (Bandura, 2006, 2008). Human behavior is not based upon any one property or factor, but is based on an interplay of determinates—intrapersonal, behavioral and environmental, mediated by the core properties. They form the core of the individual's role in self-determination and efficacy.

Intentionality. Individuals are capable of planning actions and strategies: they can be agents of change. People intend to do things, but because there are always mitigating factors that can come into play, absolute agency does not exist. Collective intentionality is required for achievement of goals by groups, with the additional factor of coordination required.

Forethought. Bandura calls forethought "the temporal extension of agency" and "anticipatory self-guidance" (Bandura, 2006, p. 164). Forethought gives intentionality direction. Goals are planned for future achievement, and likely outcomes are projected that can guide actions. Purpose can be determined. Forethought requires cognition in representation or visualization of a future that does not yet exist, but can act as motivation. "A forethoughtful perspective provides direction, coherence, and meaning to one's life" (p. 165).

Self-reactiveness. Self-reactiveness is the ability to turn the plans and goals generated in forethought into actions. It involves making choices, taking initiative, and reaching conclusions.

Self-reflectiveness. Self-reflectiveness is the metacognitive ability that only humans possess to reflect upon one's own thoughts and actions and interpret them. The individual's reflective conclusions are then used as the basis for judging those actions and for planning future actions based on those judgments. Humans judge their own self-efficacy in thought and action. This theory has particular relevance to the study of strategic test-taking.

Personal agency. Social Cognitive Theory addressed the element that was omitted from many of the previous behaviorist theories that addressed human behavior: the independent intervention by the human subject, an expression of will that is working, or *agency*. Bandura saw agency as a function that interacts with, but is independent from, the environment. "To be an agent is to influence intentionally one's functioning and life circumstances" (Bandura, 2006, p. 164). In another resource, Bandura ended the same definition with "environmental events," rather than "life circumstances"(Bandura, 2008,

p. 87). "Broadly speaking, *agency* is the capability of individual human beings to make choices and to act on these choices in ways that make a difference in their lives" (Martin, 2004, p. 135). In Martin's sense, agency is always personal.

Bandura chose to examine the idea of agency in terms of interactions at various distances from the self. He differentiated between three modes of agency: (1) personal agency, which is carried out by an individual; (2) proxy agency, in which the individual uses personal influence to motivate others to initiate action that benefits them; and (3) collective agency, in which people form groups in order to reach a mutual goal. Social Cognitive Theory explains the way humans use their own personal agency to develop, adapt and make changes (Bandura, 2002). Personal agency is the most proximal: people use it to influence their immediate environment and events in which they participate (Bandura, 2006). Bandura believed that the personal agency is the most significant type of agency, and that the most important factor in personal agency is *personal efficacy*, which is also referred to in the literature as *self-efficacy*.

Spiritual agency. Much of the self-efficacy research has been conducted to address specific psychological and/or physiological constructs that lend themselves to experimental designs. A conspicuous absence is the existence of a spiritual component in the agency literature as a whole, and in research studies in particular. One large, complex and rigorously designed study did, however, include the spiritual or religious aspect of the student academic experience as a research factor within the context of coping.

Spiritual agency in testing. In a complex and multi-faceted research design conducted in Israel, Zeidner examined multiple ways a sample of 241 college students coped with examination stress, including the spiritual component. Predictor variables

were developed from the information gathered from administration of the Spielberger's Text Anxiety Inventory and Hammer and Marting's Coping Resources Inventory. Selected items from Carver, Scheier and Weintraub's COPE Scale were used to assess the subjects' coping strategies (Zeidner, 1995).

Zeidner identified and defined three specific and discrete types of coping strategies: (1) problem-focused, (2) emotion-focused, and (3) avoidance (1995). Problem-focused strategies solve the problem, or at least manage it, by addressing the stressor and removing or getting around it. Emotion-focused strategies do not address the stressor, but address the individual's response to the stressor. Avoidance strategies are further differentiated into two types: (1) those that involve others, perhaps as distracters, and those that involve tasks, such as unnecessary activities that support procrastination. Although not specified, but consistent with this study, prayer might be considered an example of seeking out of others, and therefore, an avoidance strategy.

One aspect of Zeidner's design examined the influence of two key Criterion Variables and 11 Predictor Variables on student test outcomes. One of the key Predictor Variables was *Spiritual*. Five coping resources scales were also compared: cognitive, social, emotional, *spiritual*, and physical (Zeidner, 1995). A regression analysis showed that "only spiritual resources...made an independent contribution to the regression....Students higher in spiritual coping resources tended to engage more often in avoidance strategies than those with poorer spiritual resources" (p. 289). The least-frequently-used examined strategies reported by the 241 subjects were alcohol and drugs, and the second-least-frequently-used were denial and religion.

Zeidner (1995) found that personal resources did predict coping strategies, but the correlation was weak. “Cognitive resources are inversely related to emotion-focused strategies...whereas emotional and spiritual resources are positively related to emotion-focused coping” (p. 294). One must note that only a slight change in the definition of the terms results in a sea change in meaning. There is a relationship between emotion and spirituality, but it is not necessarily a causal, negative relationship. Using emotion-based coping is considered less desirable than using one’s cognitive resources. Spirituality should not be considered an emotion, however, but rather a separate element related to the soul. Spirituality is not simply a cognitive activity of the mind.

An apparent conclusion might be that any strong resources should boost the test-taker’s self confidence, but Zeidner saw spiritual resources as being helpful in dealing with stress only in the context of traditions and perceptions, not directly. He viewed traditions and perceptions in much the same way a Christian scholar might view a Biblical worldview—something that helps one interpret the meaning of, and decide how to cope with, stressors. The subjects in Zeidner’s study exhibited the recognition of spirituality as a resource, but generally did not demonstrate intentional agency to integrate it in a meaningful way to the task of test-taking.

Bandura’s fortuitous determinants. Bandura acknowledged that “psychological theories have neglected the fundamental issue of what determines people’s life paths,” (Bandura, 1982, p. 747), and instead focused on how people thought and behaved. The consensus of psychologists, based on Freud’s ideas, had been that the events of childhood laid the foundation and formed the framework for who the adult person would eventually become. Stage theorists, such as Piaget, saw the individual’s entire development as

predetermined by the childhood experiences. Theologians, in contrast, wrestled with the conflict between absolute agency of God, and how much agency humans actually had in exercising free will (Bandura, 2008). Bandura rejected both the nonteleological viewpoint of evolution and the theological construct of a Creator Who possesses absolute omnipotence, and attempted to define a middle ground that tended to be more humanistic (2008).

Bandura noted that "through their actions people create as well as select environments" (1982, p. 747) which is congruent with the Agency Theory. His idea of creating one's environment was also based on *chance*: he argued that "chance encounters play a prominent role in shaping the course of human lives" (Bandura, 1982, p. 748). He credited many of the actions humans take to experiencing a *chance event*, which he defined as "an unintended meeting of persons unfamiliar with each other" (Bandura, 1998, p. 95). When series of events intersect, Bandura considers that a *fortuitous encounter* (Bandura, 1982).

In spite of his large body of published experimental research on a vast range of topics, Bandura supported his contentions about chance, fortune, and spiritual issues almost exclusively with personal anecdotes and observations in which he reflected about individuals' vulnerability to significant life change based on chance encounters (Bandura, 1982, 1998, 2008). Bandura credited fortuitousness with the play of events, not the will or design of a Higher Power, and definitely not God. He elaborated extensively on how an individual person's own agentic activity can manage fortuity, which is a decidedly humanistic approach.

Bandura did address the subject in a chapter he wrote for an edited text on free will, which he entitled, “Reconstrual of ‘Free Will’ from the Agentic Perspective of Social Cognitive Theory” (Bandura, 2008). In it, he again emphasized fortuity, but also addressed the concept of morality, and once again aggressively discussed and rejected the biology-based theories of behavior and learning that leave out the agentic independent volition of the person.

Biblical Christian view of agency. In contrast to a belief of good or bad fortune establishing the context for behavior, Christians acknowledge the existence of human agency within the framework of the workings of God. Ultimate truth, rather than being unknowable, or situational, is explained in the Bible and personified in Jesus Christ. Ultimate responsibility for outcomes rests with God, but the person is responsible for obedience and acceptance of God’s will.

Will. Numerous references in the Scriptures refer to the will of God (Ps. 40:8; Romans 12: 2; I Thes. 4:3, 5:18; 2 Tim. 2:26; Heb. 10:7, 13: 21, English Standard Version). The references to human agency tend to be assumption that humans have the ability to respond positively to admonitions, encouragements and imperatives to complete specific behaviors: “Let us…” (Heb. 12:28), “Believe on …” (Acts 16:31), “If you confess…” (Rom. 10:9). The ultimate demonstration of human and supernatural wills was the great expression of submission by Christ in the Garden of Gethsemane, “Not as I will, but as You will” (Mat. 26:39). In each case, however, as an action was considered, the individual who had to decide whether or not to initiate the action. The result was based on a decision, not a conditioned response. The person was the agent by whom the

action would be carried out, the one with the intention to influence. In that sense, there is some congruence with self-efficacy theory.

Within the parameters of Bandura's explanations of agency, however, self-efficacy could be fully functional within agnostic or atheistic humanism. Within the parameters of a Christian worldview, however, all agency, human and angelic, has limitations that are established by God. In the account of the fall of Lucifer (Is. 14:12-17), his repeated transgression was a series of five "I will…" statements, culminating with "I will make myself like the Most High." But beginning with verse 15, God states what actually will occur: events that demonstrate that His power supercedes Lucifer's willful boasts. New Testament doctrine supported the same principle. The Apostle James cautioned his readers not to be presumptuous in furthering the activities of their own wills: "Come, now, you who say, 'Today or tomorrow we will go into such and such a town and spend a year there and trade and make a profit'—yet you do not know what tomorrow will bring" (Jam. 4:13-14a). His directive was clear: "Instead you ought to say, 'If the Lord wills, we will live and do this or that'" (Jam 4:15).

Predestination. Human agency, from the Biblical Christian standpoint, is interpreted through more than one school of thought, based primarily on the difference in interpretations of *election*, *predestination* and the *foreknowledge* of God. "'Predestine' means to mark out or determine before-hand" (Note for Ephesians 1:11, ESV). Being omniscient, God knows all that will transpire as a result of human choices before they occur, and gives the individual the freedom to make the choices. Since He knows it, it will come to pass. Bandura specifically rejected the idea that life events can happen by design, instead of fortuitously, to provide opportunities for agentic responses. He stated

that "nonagentic conceptions strip humans of agentic capabilities, a functional consciousness, and a personal identity" (2008, p. 104).

Consequences. Within the Biblical Christian worldview, humans are responsible for the choices they make. One somewhat grim illustration is Jesus' pronouncement about the role of Judas Iscariot as His betrayer, in Luke 22:22: "For the Son of Man goes as it has been determined, but woe to that man by whom he is betrayed!" Another is Acts 2:23: "This Jesus, delivered up according to the definite plan and foreknowledge of God, you crucified and killed by the hands of lawless men."

Clearly, individuals do have choices: "Choose this day whom you will serve…but as for me and my house, we will serve the Lord" (Josh. 1:8); and just as clearly, there are consequences of the choices: "For the wages of sin is death, but the free gift of God is eternal life in Christ Jesus our Lord" (Rom. 6:23). The degree to which one's spiritual life is related to one's physical psychological and academic life is largely dependent on the level of agency the individual applies to that relationship, and the choices that are made.

The stress or balance between personal agency and God's will is evident in the educational setting, and particularly in the area of assessment. A student might say, "I read the chapters once, I listened in class, and I did my assignments. I'm not going to go back and study or review anymore. Whether I pass the test or not is up to God." Faculty and conscientious students, however, would more likely say, "I read the chapters once, I listened in class, and I did my assignments. Now I'm going to ask God to bless my study and review time for the test so I will understand and remember what I have learned." There is a degree of personal agency involved.

Self-Efficacy Theory

The model. Bandura defined self-efficacy as "people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives" (Bandura, 1994, p. 1). The three interrelated components of self-efficacy are the (1) Personal Determinants (the individual), (2) Behavioral Determinants (the behavior, i.e. actions and reactions of the individual), and (3) Environmental Determinants (the environment or context) (Bandura, 2008). The interrelationship of the three factors resulted in what Bandura termed Triadic Reciprocal Determination, a triadic model, in contrast to the earlier dyadic model developed by Sears. Personal Determinants, the additional element Bandura added to the dyad, comes from within the individual and is exhibited as *will*, *agency*, or *self-efficacy*.

The individual became, in Bandura's theory, a separate element with self-determination, not simply a responder to a stimulus. Bandura believed that all the theoretical elements and their expressions have mutual influences upon each other. Modeling, teaching, and social pressures impact the individual, who can respond by impacting or leaving the environment, which then changes the way the environment can subsequently impact the individual. This triadic view supports the proposition that students can exert influence on their own testing experience through agency.

Domains. According to Self-Efficacy Theory, "people develop domain-specific beliefs about their own abilities and characteristics that guide their behavior by determining what they try to achieve and how much effort they put into their performance in that particular situation or domain" (Grusec, 2009, p. 782). More specifically, Bandura saw that the beliefs people hold about their own capacity to develop motivation is a

critical element in their ability to complete the action successfully (Bandura, 1994; Pajares, 2003). How effective they perceive themselves to be influences how effective they are. In a large study comparing individuals' action plans with the completion of the plan, Lippke, Wiedemann, Ziegelmann and Schwarzer (2009) concluded that a higher degree of perceived self-efficacy translates into a higher ability to act upon one's intentions.

The Self-Efficacy Theory aspect of Bandura's Social Cognitive Theory has been applied to research in a number of academic domains, including academic achievement. Regardless of other motivators at play in a given situation, the individual needs to believe that his or her actions can produce a desired effect or outcome, or there would be no point in initiating the action. Bandura specified that human behaviors are impacted by the individual's self-efficacy belief "through cognitive, motivational affective and decisional processes" (Bandura, 2002, p. 270). Numerous studies have demonstrated support for these constructs, and recently there has been particular emphasis in their application to formal educational settings (Martin, 2004).

Sources of self-efficacy. Bandura (1994) identified four broad sources of, or influences upon, an individual's self-efficacy: (1) mastery experiences, (2) vicarious experiences, (3) social persuasion, and (4) interventions which "reduce people's stress reactions and alter their negative emotional proclivities and isinterpretations [*sic*]of their physical states" (p. 3). First, when an individual survives adversity and setbacks to master a task, the success provides support for the belief that he or she can succeed again. As that belief strengthens, resiliency and perseverance result, in addition to a perception of self-efficacy. Second, observing and learning from the successes and failures of others

allows the individual to vicariously gain useful knowledge and efficacy without the consequences of trial and error learning strategies. Third, verbal encouragement often promotes an increased will to try harder, which can result in increased efficacy. Fourth, learning to re-interpret one's perceptions of stress reactions, moods, and emotional states in the context of a challenge from negative to positive improves one's perceived self-efficacy. People's belief in their ability to cope affects their ability to cope.

Influences of and Responses to Bandura's Theories

Although Albert Bandura has continued to write and theorize into the 21st century, numerous others have taken his earlier work and enlarged the theoretical frameworks it established through research. Some have accepted the challenge of developing measurement instruments. Still others have responded by challenging the veracity of Bandura's ideas.

Model development. Miller and Brickman (2004) developed a model based on Bandura's concepts of proximal and distal goals. They believed that distal goals act as motivators in formulating proximal sub-goals which support achievement of the distal goals. They supported their theoretical model with an integrated literature review, rather than research.

Quantitative research. Longo and Lent (1992) were concerned with the attrition rate of counseling and psychotherapy patients, and expressed regret at the lack of research that provided any answers. Their study attempted to interpret and apply Bandura's Social Cognitive Theory to the problem of attrition and see if the link were viable. The outcomes expectations aspect was of particular interest to them. Both attrition and outcomes expectations are important in educational research. Longo and

Lent's subjects were 139 university students who had voluntarily sought the services of a counseling center. The instrument was a Likert-like assessment tool and several copyrighted questionnaires. They found non-significant correlations between self-efficacy and self-esteem; but self-efficacy, outcome expectations and motivation showed significant positive correlations. Outcome expectations and motivation are important factors in test-taking.

In the Longo and Lent study, a statistical equation developed for gender, problem distress, outcome expectations and self-efficacy showed significant, unique predictive variance, and 47% of the variance in motivation could be accounted for by the full set of variables in the equation. They concluded that "the findings of this study suggest that social cognitive theory may help explain clients' motivation for, and perseverance in, counseling. Both self-efficacy and outcome expectations accounted for significant unique variance in motivation…" (p. 7). This study might have generalizability to learning of test-taking strategies, due to the similarity in the nature of learning an academic coping strategy and learning psychological coping strategies. Further research is needed.

Self-efficacy assessments. Bandura formalized his Self-Efficacy Theory in 1974. Since then, the concept has resonated in numerous disciplines and has been extensively researched. A search of the social sciences research database SocINDEX with Full Text, using "self efficacy" as the keywords, produced 2,057 journal and book titles containing 141,505 articles.

An advanced search of the SocINDEX with Full Text database using a Boolean operator with keywords "self efficacy" AND "assessments" and limited to journal articles

published since 2000, produced 18,847 articles. Adding “tools” as a keyword, in order to find assessment tools, reduced the results to 4,610 articles. The addition of “instrument,” as a keyword still resulted in 2,350 articles. Bandura’s Self-Efficacy Theory has had a profound influence on educational research.

The General Self-Efficacy Scale. Matthias Jerusalem and Ralf Schwarzer (1995), authors of studies on stress and coping and colleagues at the Free University of Berlin, were co-authors of the General Self-Efficacy Scale (GSE), a generic self-efficacy assessment instrument. The date of development was listed by one of the authors, Schwarzer, as 1979 in one primary source (Schwarzer, n.d.) and as 1981 in another (Schwarzer, 2009). Dates vary in other sources, as well, but may reflect the date the instrument was translated into the language used for a researcher’s particular study. There is also variation in use of representative acronyms: GSE (Schwarzer, 2009), or GSES (Statistics Solutions, 2009). The acronym differences may be attributed to translation issues.

The GSE has been widely used and extensively studied throughout the world. At the 2003 annual convention of the American Psychological Association, Luszczynska (2003) reported a meta-analysis of published validation studies using the GSE. Her analysis identified a total of 19,120 research participants in 25 countries. She isolated specific factors and examined them as a five-country subset consisting of 8,796 participants in Costa Rica, Germany, Poland, Turkey and the United States. She concluded that “correlations between the GSE scale and other measures of personality factors, such as optimism, self-regulation, orientation towards the future, and self-esteem, remained significant and mostly in the moderate range” (p. 1). Validity was also

examined in studies with subjects as diverse as job burnout, health behaviors, and online data collection. She concluded that general self-efficacy is a universal construct.

Two years later, Luszczynska, Scholz and Schwarzer (2005) published a meta-analysis of GSE validation studies with 1,933 participants in Germany, Poland and South Korea. They identified four sets of variables that exhibited positive associations with self-efficacy and again concluded that "general self-efficacy appears to be a universal construct that yields meaningful relations with other psychological constructs" (p. 439).

Self-efficacy and the college student. Bandura (1986) described self-efficacy as a *mediating mechanism* of personal agency. He saw it as mediating between prior influences and subsequent behavior. The field of educational psychology has moved toward a focus on seeing students' self-efficacy beliefs as a critical component in their ability to be successful in their educational settings (Pajares, 2003). Woolfolk, Winne and Perry (2000) wrote that educational psychologists seek to understand "a combination of academic learning skills and self-control that makes learning easier, so [learners] are more motivated, in other words, they have the skill and will to learn" (p. 384). First-time college attendance is often the highest level of independence an individual has experienced to that point. The student's prior influences through 18 years of rearing and 12 years of formal education are bridged to academic and personal success in the transitional environment of college partially by their self-efficacy beliefs and how they operationalize them. Effectiveness and success can be defined differently from individual to individual, but there is, in each case, the need to perceive the ability to achieve as an individual, a self. The self must have the will to make the attempt to be effective.

Self-efficacy and student affairs. Self-advocacy as an expression of self-efficacy has been studied in higher educational settings with very specific populations. DiRamio and Payne (2007) designed a self-efficacy study in student affairs with a population of approximately 9,600 undergraduate university students. They used an attitude survey (N=888) to assess opinions about available services offered in student affairs to better inform the institution's offerings. Results were treated statistically for *t*-scores and effect size. The study had a strong design. The limitations were largely based on the homogeneous nature of the sample, although the sample was so large, there may still be valid generalizability of the range of statistically significant results.

Critics of Bandura's theories. Throughout the decades since Bandura began formulating and articulating his theories, his ideas have had critics as well as supporters. Critics challenged the authenticity of the concepts he promulgated as well as his understanding of the scientific foundations of them (Powers, 1991; Rottschaefer, 1991). Bandura responded with a lengthy, lucid but complex article that attempted to answer the criticisms (Bandura, 1991). Others charged that he oversimplified complex factors involved in behavioral change, or that his causal factor understandings were flawed, charges he also refuted (Pajares, 2004). Even if all of the accusations were valid, the outcomes of the hundreds of studies based on his theories and the ongoing development of new theoretical models for which his theories were foundational would appear to provide strong conformational evidence for their validity and reliability in practice.

Stress, Anxiety, Coping and the Test-taker

Bandura described agency as " to influence intentionally one's functioning and environmental events," (Bandura, 2008, p. 87). "Broadly speaking, *agency* is the

capability of individual human beings to make choices and to act on these choices in ways that make a difference in their lives" (Martin, 2004, p. 135). Some life events are not easily influenced, and the attempt to exert influence creates various, and sometimes stressful, responses.

Definitions

Definitions of stress differ greatly from one academic discipline to another. Stimulus-response theorists see stress as any stimulus. Lazarus and Folkman discussed how a stimulus acts upon the individual (1984). Sides, writing from the medical/physiological perspective, defined stress as "simply the wear and tear imposed on the body by the normal or abnormal events of life" (1998, p. 59), taking a more environmental than active approach. Even theorists who largely agree on the definition, however, differ on whether stress can be defined to include being either positive or negative, or only negative.

For their research, Lazarus and Folkman (1984) accepted Elliott and Eisendorfer's 1982 definitions of the four kinds of stressors: (1) acute, time-limited stressors, (2) stressor sequences, (3) chronic intermittent stressors, and (4) chronic stressors (p. 14). College test-taking would be considered a chronic intermittent stressor. It is ongoing, but not continual.

Individuals react differently to the same stressors, but the responses reflect adaptation and coping. Coping has been defined as "the behavioral and cognitive processes used when individuals are attempting to deal with the demands of a stressful situation" (Thompson & Gaudreau, 2008, p. 269), and as "constantly changing cognitive and behavioral efforts to manage specific external and/or internal demands that are

appraised as taxing or exceeding the resources of the person" (Lazarus & Folkman, 1984, p. 141). Bolger stated that coping is "a process explanation for differences in stress outcomes" (1990, p. 525). Coping mechanisms can range from active confrontation and engagement to total withdrawal from the stressor. The individual uses cognitive appraisal to evaluate the situation and decide how to respond, but involuntary psychological or emotional responses can override one's intentions.

Taking a test--being assessed or evaluated based on one's knowledge has been viewed in the research as a stressful situation (Cohen, Ben-Zur & Rosenfeld, 2008). High-stakes tests, by definition, have significant impact on the test-taker's life, and therefore present the opportunity for a stress reaction, which is often cognitive anxiety—the subject knows he or she is stressed. A great deal of research has been conducted to examine the links between stress and test-anxiety, focusing on cause/effect relationships, and the mediating elements of voluntary and involuntary coping strategies.

Stress in Testing

Swartz (2006) designed a study to determine graduate students' perceptions of the use and accuracy of three types of assessments, including multiple-choice questions, and to determine which format for multiple-choice questions provided a higher correlation with students' self-reported mastery of the content—a reflection of their perceived self-efficacy. He utilized a survey and an extensive statistical analysis of the results. He found that the subjects strongly preferred constructed response questions to multiple choice questions and confidence level assessments, while there was no statistically significant difference between the preferences for confidence level assessments and constructed response formats. Swartz's study supports the notion that college students

perceive multiple choice questions as particularly challenging. A perceived challenge can serve as a motivator to become agentic regarding test-taking skills.

Trait and State Anxiety

Coping theorists differentiate between two broad categories or dispositions of anxiety that require coping: trait and state. "Trait anxiety is the disposition to show anxiety across diverse situations. State anxiety, in contrast, is the anxiety people experience on a moment-by-moment basis" (Bolger, 1990, p. 525). Test anxiety can exhibit manifestations of both dispositions.

Zeidner (1995) noted that educational assessment provides an appropriate context for research in anxiety and coping, because it contains representative elements of significant environmental stressors that are also found in other contexts: preparation, confrontation, uncertainty, and coping. Zeidner referred to the assessments as "stressful social evaluative encounters," (p. 281), applying Bandura's recognition of the social aspect of the activity.

A meta-analysis indicated that in the dispositional model, task-oriented coping has been linked with optimism, while disengagement-oriented coping is linked with pessimism (Thompson & Gaudreau, 2008). "The generalized positive self-beliefs rooted in optimism are volitional, and their agentic function is likely to promote a sense of autonomy and competence needed to engage in activities out of self-determined reasons" (p. 272). Thompson and Gaudreau asserted that the individual who can develop task-oriented coping strategies increases intrinsic motivation. Disengagement- and emotion-oriented models, in contrast, do not motivate.

Davis and his colleagues believed there is another dichotomy in test anxiety: agency and testing problem efficacy (Davis, DeStefano & Schutz, 2008). In the context of test-taking, "agency is defined as the extent to which students appraise the outcome of a particular test as being under their control" (p. 944). "Testing problem efficacy is defined as the judgments students make about their ability to manage the problems that emerge during tests" (p. 944). The choice of coping strategy at any given testing situation can depend upon the student's repertoire of knowledge and strategic approaches to the problem (Zeidner, 2007). Zeidner reported that test-taking anxiety accounts for about 4% of the statistical performance variance in a range of assessment environments. Effectively dealing with anxiety is of extreme importance in the environment of high-stakes testing.

Coping by the Test-Taker

Coping can represent choices. Broekaerts (1993) viewed the making choices as choosing pathways. Choosing a *growth pathway* indicates perception of the stressor as a challenge with the potential for growth: this option is optimistic. A student for whom test-taking is a threat, in contrast, might choose a *well-being pathway*. Choosing a well-being pathway is more conservative and safer than opting for the riskier growth pathway, and can be implemented with less perception or acknowledgement of threat. Its highest value is achieving or maintaining a sense of well-being: it could be called the path of least resistance. One would expect less possibility for growth if a well-being pathway is chosen. "When an important goal is at stake, and students detect a discrepancy between the task demands and their personal resources to meet them, the task will have consequences for well-being" (Boekaerts, 1993, p. 154).

Davis, DeStefano and Schutz (2008) designed a qualitative study of first-year college student test-takers, and a subsequent replication of it, each with more than 1000 subjects. They identified classifications of students based on how they appraise and manage or cope with problems associated with test-taking. The study results supported Boekaert's (1993) contention that students tend to choose their test-taking approach by selecting either a growth pathway or a well-being pathway. If a student is faced with an academic task, such as a test, and perceives the task as congruent with his/her long-term goals, the student is more likely to choose the growth pathway, even though it may require work and sacrifice. It is goal directed. Choosing the growth pathway allows the student to focus, or refocus, on his or her goals.

If the task is perceived as not congruent with the student's goals, the test-taker is more likely to choose the well-being pathway, which attempts to achieve or maintain an equilibrium characterized by a sense of well-being. Approaches typical of the well-being pathway are (1) avoiding or emotionally distancing oneself from the task rather than intentionally putting oneself into an acute sense of anxiety, (2) denying the importance, urgency or consequences of the task, and (3) blaming extraneous variables for unwanted outcomes. The amount of stress the student can or is willing to tolerate is a key factor in the pathway choice made in any given testing situation, and this choice is critical in the student's success. Since the need to take a test is a chronic, intermittent stressor, the opportunity to make the pathway decision presents again and again. To maintain the highest test validity, student stress should be addressed.

Sense of coherence (SOC). Cohen, Ben-Zur and Rosenfeld (2008) studied three assessment factors exhibited by first-year college students --sense of coherence (SOC),

coping strategies, and test performance. Their participants included nursing students. The SOC is a person's internal resource "that facilitates efficient coping with negative life events and life stressors, buffering their negative effect on psychological well-being" (Cohen, Ben-Zur & Rosenfeld, 2008, p 291). Bandura's concept of self-efficacy is congruent to the test-taker's perception of being able to cope, or manage. A perception that a task is manageable indicates that the individual believes s/he has resources available that are adequate for the demand (Cohen, Ben-Zur & Rosenfeld, 2008). A test-taking strategy can become a personal resource to meet the demands of taking numerous high-stakes tests, and make then manageable. This is manifested as perceived self-efficacy.

Self-efficacy and coping. The approach chosen to use in taking a multiple choice test is a very individual matter, as is the choice of coping strategy to deal with getting through the test successfully. Renner, Spivak, Kwon and Schwarzer found that individuals with highly developed coping skills and high levels of self-efficacy not only can initiate action, but once they begin a task, they invest more effort and will attempt to persist in it longer than those with lower coping and self-efficacy skills (2007). Successful completion and achievement of a passing score on a high-stakes test often involves personal, proximal and cooperative agency.

High Stakes Tests

No Child Left Behind

High-stakes tests came into the public awareness at a national level with the passage of the No Child Left Behind (NCLB) legislation written in 2001 and signed into law on January 8, 2002, by President George Bush (Illinois State Board of Education, n.

d.). A vast infusion of recent literature on testing as an individual and institutional assessment construct is focused on various aspects of NCLB. The broadest federally-mandated assessment program ever initiated, states participated in the National Assessment of Educational Progress (NAEP) for assessment to determine if their defined adequate yearly progress (AYP) had been met by each student. The penalties for schools whose students did not meet the criteria began with special designations such as *low performing*, progress through various sanctions, and could be as severe as complete loss of local autonomy. The NCLB tests were, therefore, extremely high-stakes, and the risks so great that ordinary citizens became aware of them through the pervasive impact the requirements had on local public schools.

Some of the most significant results of the NCLB legislation have been the plethora of accountability assessments nationwide, preparation for them, and analysis of them (Judson, 2007). The mandates of the NCLB legislation have spurred research into what kinds of preparation for high-stakes tests works best, and what kind of remediation is most successful. Much of it is intended to be generalizable to any level of education, including higher education. Test-taking has become a skill that is important not only to the individual, but has become significant at the institutional level.

Challenges. There are challenges in designing research that assesses test-taking strategies. Cohen (2006) noted that there is an inherent difficulty in assessing test-taking strategies without intruding into the process. Researchers have typically relied upon verbal reports from the test-takers. Those reports could be self-reports, self-observations, or self-revelations (Cohen, 2006). In the 15 years preceding Cohen's study, however, new emphases developed that involved test validity and its relationship to test-taking, the

relationship of the test-taker's language proficiency to the test outcomes, and the effectiveness of teaching test-taking strategies. He concluded that more well-designed, rigorously-conducted research was needed.

Limitations in the Literature

There is a significant body of literature that addresses the anxiety and stress associated with test-taking. The emphasis in both theoretical and practical applications is a psychological approach. Only a few published studies addressed the actual activity of taking a test, and they each had specific components that limited their generalizability.

Second language students. In 2006, Andrew Cohen of the University of Minnesota wrote a summary of the published research on test-taking strategies of the previous 25 years. He had published a previous one in 1984. He noted that "what was missing was the aspect of test validation that related to respondents' behaviors in taking the tests; little was known about what they were actually doing to produce answers to questions and how this corresponded to the abilities one sought to test" (p. 307). The main limitation to Cohen's study was that although it was not specified to be exclusive to second language students, it was largely specific to them, nevertheless. Four of the five conceptual frameworks he identified were specific to English language learners, and eight of his references were to his own published articles. That peculiarity in the sample may limit generalizability.

All self-reported data. Rachal, Daigle and Rachal (2007) studied college students' self-reported strengths and weaknesses in learning and assessment strategies, looking for differences based on class level in college. In every academic level classification, more than 50% of the students reported difficulties with test-taking.

Because test-taking is ubiquitous in college settings, college students are always anticipating a test, taking a test, or considering a test already completed. All three positions relative to tests—before, during and after, are open to stress, but are also opportunities for appraisal and action. The limitation to generalizability is the necessary subjectivity of relying on individuals' own self-reported post-activity perceptions.

Focus on psychological stress. The way an individual deals with a stressor is by coping with it. One's proficiency at coping is a reflection of one's self-efficacy. In test-taking, there is a need for coping with anxiety before, during and after the stress of the actual test-taking experience. That anxiety can be reduced, or mediated by strategic actions. The longer the anticipation time, the more potential there is for complexity in appraisal because of mediating coping processes. Given time, people can reflect, suffer or grieve; they can also avoid the problem, think about it, take action or make efforts to gain self-control. Each of these intervening coping processes will affect subsequent appraisals and their accompanying emotions. (Lazarus & Folkman, 1984, p. 98)

There is clearly a need for further objective experimental research about what actually happens during a testing situation.

High-Stakes Tests in Nursing Education

Nursing students are confronted with high-stakes, multiple-choice format tests as assessments of their core knowledge at all levels of their pre-professional education. Even after they complete the rigorous curriculum required to achieve a college degree in nursing, they cannot become practitioners until they pass another high-stakes multiple choice test, the NCLEX. Throughout their programs, they take literally thousands of

multiple-choice questions, both as assessments and as practice for higher-stakes assessments.

Test banks. A pivotal work on multiple-choice questions for nursing students analyzed the questions themselves. Six advance-practice nursing instructors from the Lansing School of Nursing and Health Sciences, Bellarmine University, Louisville, Kentucky, completed a comprehensive research study of the multiple-choice questions in 17 test banks accompanying nursing textbooks with a total of 2913 questions (Masters, Hulsmeyer, Pike, Leichty, Miller & Verst, 2001). Table 1 summarizes their findings.

Table 1

Multiple-Choice Questions in Nursing Test Banks (Masters et al, 2001)

Research Factors	Findings
1. Adherence to widely accepted test-writing guidelines	2233 violations in 2913 questions
2. Bloom's cognitive level	Percentage found at each level: 6.5% Analysis 21.8% Application 24.8% Comprehension 47.3% Knowledge
3. Distribution of correct answers	Fairly even spread between letters A, B, C, D: range 22.8% to 29.6%

Masters et al developed an evaluation instrument of 30 guidelines taken from earlier literature they used to evaluation the test bank questions. One limitation noted was that no maternity nursing text was included, because none of the researchers had adequate advanced expertise in that specialty to adequately evaluate the questions and detect small content errors or omissions.

The ratio between the number of errors and the number of questions was approximately 3:4. More than twice as many questions were written at the most basic Knowledge level than at the critical thinking Analysis level. Five of the 17 test banks had no questions at the Analysis level. The authors of the study interpreted their findings as strong evidence that poorly-constructed multiple-choice test questions are in general use in nursing education and assessment.

Instructor-written questions. While numerous test banks of questions are available in nursing education, Su, Osisek, Montgomery and Pellar (2009) reported a locally-designed and implemented program to increase nursing faculty's ability to design their own multiple-choice test questions. They based their approach on recommendations by Morrison, founder of HESI:

(1) Design questions that reflect application or analysis cognitive level.

(2) Require students to apply knowledge of more than 1 concept to answer the question.

(3) Require a high level of discriminating judgment by asking students to choose the best/most appropriate option from all plausible alternatives (p. 223).

This advice assumes the questions were written to Bloom's *Analysis* level.

Su, Osisek, Montgomery and Pellar (2009) offered a variety of advice for new nursing faculty learning to write tests: When writing distractors, base them on common misconceptions, write them in syntax that has the appearance of correctness, and design "homogeneous alternatives that reflect parallel content" (p. 225). If a new instructor has insufficient formal background in education to write tests well, it is questionable that such recommendations could be understood and utilized.

HESI. In addition to instructor-written tests and proprietary test banks, standardized high-stakes tests that are intended to predict the statistical probability of passing the NCLEX are widely used in nursing education. Health Education Systems, Inc. (HESI), was begun by Susan Morrison, PhD, RN, in the early 1990s to provide nursing assessments (Evolve: Student site: Evolve reach, 2008). HESI was purchased by international publishing conglomerate Elsevier. In a January 9, 2006, press release in Amsterdam, Netherlands, Elsevier stated that "a primary objective of HESI products is to provide faculty with the tools to predict how their students will perform on the NCLEX exam, a licensure exam for nursing students" (Reed Elsevier, 2006). Morrison, who was retained as a corporate officer, emphasized the predictability factor for HESI when she stated, "HESI products are recognized by nurses in both academic and clinical practice settings as having exceptional outcome predictability, a feature that is highly valued by our clients" (Reed Elsevier, 2008).

Use. HESI examinations are administered only under secure situations by qualified administrators, either on college campuses or online, for a fee. Elsevier's HESI division offers statistical analysis of a student's HESI test scores relative to other students in the same program and in other similar programs, as well as detailed sub-scores in

numerous areas of their Admission, Exit, Specialty and Custom Exams (Evolve Instructor Site, 2008). Most are designed to be formative relative to a given level of a specific body of knowledge, such as obstetrical nursing, or medical-surgical nursing; but the HESI Exit Examination (E^2) is designed to be administered near or at the end of the nursing program, and is presented as a statistical predictor of success on the NCLEX. According to HESI database records, the number of schools of nursing using HESI exams increased from fewer than 90 in 1999 to more than 560 in 2003 (Morrison et al, 2004).

In a peer-reviewed journal article, Ainslie T. Nibert, PhD, RN, formerly Director of Research for HESI, and under the Elsevier ownership, Vice President, wrote with two collaborators, and connected the HESI high-stakes tests with predictability and remediation:

> A new and growing trend among nursing faculties is to set benchmarks that contain progression-to-graduation requirements based on student performance on standardized nursing exams that allow for comparison with national norms. These policies are designed to identify students who are in need of remediation prior to graduation and NCLEX candidacy so that remediation can be initiated and NCLEX failure averted. (Nibert, Young & Britt, 2003).

High stakes. The most critical piece of the reporting of the HESI scores is the predictor score: the percentage of chance of passing the NCLEX on the first attempt. HESI recommends various cut-off points for progression and graduation based on these predictor scores and a significant number of programs have adopted those recommendations as policy (Nibert, Young & Britt, 2003; Spurlock, 2006). There are high stakes involved: according to a survey of programs using the HESI, the most

common sanctions imposed for failure to reach the required score include inability to progress to the next semester, inability to graduate, and inability to sit for the NCLEX (Nibert, Young & Britt, 2003). Imposing even one of those consequences makes the HESI tests high-stakes.

Concerns. At least four validity studies of HESI tests have been conducted, and they have served as the basis for assurances that nursing programs can legitimately place sufficiently high levels of confidence in the predictability scores to use them as elements in progression policies—that is, as gatekeepers to prevent progression or graduation. The third (Nibert & Young, 2001), and fourth (Nibert, Young & Adamson, 2002) studies were published in a peer-reviewed nursing research journal, but the primary researchers were people directly affiliated with HESI, rather than neutral evaluators. Elsevier's purchase of Mosby, the parent publisher of the Mosby AssessTest, and the removal of that competition from access as a secure test, almost certainly significantly contributed to HESI's subsequent exponential growth, as well as its insularity. Now Kaplan is HESI's only major competitor in providing standardized NCLEX predictor assessments.

The statistical applications used for the HESI predictability have been given a proprietary name, the HESI Predictability Model (HPM) (Nibert, Young & Britt, 2003), and are a closely-held secret. The level of secrecy HESI maintains regarding its statistical applications that would demonstrate the validity and reliability of their high-stakes tests has raised some concerns (Downey, 2006). Spurlock noted that a 1999 publication of the American Educational Research Association suggested that test-use validity should be strengthened by allowing external experts to review the specifications and procedures, and validate the statistical methodology used for scoring (2006). Like

Downey, Spurlock also expressed concern that HESI has not permitted either of these to be done. He also noted that he was also concerned that Morrision, with Adams and Hsia (2004) stated that there was still a need for studies of the item functions on the tests.

Conclusions. Questions about reliability and validity of the high-stakes tests make the ethical and practical challenges of faculty who must prepare their students to pass, and remediate them when they do not, even more challenging. How does one teach students to prepare to answer examination questions for which the internal or construct validity cannot be determined with certainty? Maximizing their analytical test-taking abilities to focus their attention on the content being assessed instead of the questionable elements of the question itself is one option.

Institutional Responses to Nursing Students' Testing Needs

Nursing programs are highly invested in the academic success of their students. The data obtained from administering predictive tools like the HESI informs both curriculum and pedagogy. Faculty attempt to select assessment instruments that are the best predictors of NCLEX success. They develop policies and procedures for maximizing benefits from them. They establish institutional practices for taking high stakes tests at times in the academic year that will afford maximum opportunity for preparation and remediation, so there is adequate time for students who are unsuccessful on their first attempt to prepare and retake the assessment before the time constraint of the next academic year or graduation.

Best practices. Frith, Sewell and Clark (2005) proposed what they considered to be appropriate NCLEX preparation in an article entitled, "Best Practices in NCLEX-RN Readiness Preparation for Baccalaureate Student Success." Their strong instructional

focus of both preparation and remediation is based on study and review of HESI's multiple choice questions. The interpersonal component is consistent with social learning theory.

High-stakes tests and remediation. With the severely limited number of options for predictability of first-time NCLEX passage, nursing faculty have looked closely at the HESI examinations in determining best practices in their usage and remediation following failures. Colleges have a range of available options relative to students who do not pass required high-stakes tests. The college can simply notify those students who were unsuccessful on the first attempt, and announce when the next test session will be, leaving remediation up to the students. At the opposite extreme, a college may mandate a highly prescriptive remediation program, and enforce it. A range of opportunity combinations exists between those extremes. Even when the school offers a range of remediation opportunities, all students who need them may not avail themselves of them. Some will choose to study independently, hire personal tutors, pay for remediation programs, or access non-secured tests for practice, and remediate on their own. Some will seek no help at all. Whichever response they choose, the stakes are high.

The American Psychological Association (APA) links high-stakes tests with remediation plans. The organization addresses the issue of high-stakes tests in an article on the APA Online website (2007), in which stakeholders in high stakes testing are warned: "Tests, when used properly, are among the most sound and objective ways to measure student performance. But, when test results are used inappropriately or as a single measure of performance, they can have unintended adverse consequences." The article cites *Standards for Educational and Psychological Testing* to list principles that

should guide any program of high-stakes testing. The authors conclude that when high-stakes tests are used, certain specific safeguards must protect the students taking the tests, specifically, programs for remediation (APA Online, p. 2). One remediation element could be mastery of an effective test-taking strategy.

The HESI Reach and Exit Examination tests are widely used as high-stakes tests that result in the need for subsequent remediation. Nibert surveyed administrators of 158 RN schools that use the Exit Examination as a "progression benchmark and remediation guide" (Nibert, Young & Britt, 2003) and reported that of the 149 that responded, 45, (30%) had implemented or continued a progression policy related to the HESI, with the following consequences of failure to meet the required score: denial of graduation (51%), incomplete or failing grade in the capstone course (34%), and withholding of approval to take the NCLEX (14%). Eighty percent required retesting, presumably following remediation, using a different version of the test.

Do the ends of high stakes test success justify the means of drastic penalties for failure on the tests? What is the faculty, and the school's, responsibility in preparing student to be successful on high stakes multiple choice tests that are an inescapable part of the nursing education programs? Nursing faculty everywhere must address those questions, and a great deal more research is certainly warranted.

Conclusions from the Literature on High-Stakes Tests

Zeidner (2007) believed that statistical variance in construct and cognitive domain validity can be negatively impacted by the test-taking abilities of the students taking the test. Since the implementation of NCLB, high-stakes testing has been of concern at all levels of education, but the research emphasis has been on learning content in a way that

will be reflected in test scores and psychological mechanisms for coping with the stress and anxiety associated with taking a high-stakes test. In nursing education, gatekeeper high-stakes tests are obligatory at multiple junctions in the curriculum, culminating with the post-graduation licensure examination, the NCLEX. Concerns about the predictive validity and gatekeeper use of the widely-used HESIs were expressed. Evidence has been published that demonstrates the potential for poorly-written test questions to be the measures of the high-stakes outcomes. Students need to have not only learning skills, but strategic test-taking skills to maximize their ability to validly represent what they know in the face of several mediating factors.

Strategic Test-Taking

Lazarus and Folkman noted that skills learned for one domain could be transferred to another, in that "problem-focused coping strategies are similar to strategies used for problem solving. As such, problem-focused efforts are often directed at defining the problem, generating alternative solutions, weighting them, and acting" (Lazarus & Folkman, 1984, p. 152) The development of test-taking strategies was consistent with the premise of building on the student's previous knowledge, such as how to problem-solve, but was not generally supported by evidence.

Test-Taking Support

A plethora of material was available on test-taking strategies, suggestions, tips, and prompts that appeared to be based on conventional wisdom or anecdotal experiences. Those that did have citations often cited other resources that were not based on research. Many of those that were based on research were published in the 1970s. The otherwise unsupported strategies seemed to be self-perpetuating. With only very rare exceptions,

the recent articles found that were not NCLB-specific addressed the needs of students with various disabilities, which was probably a reflection of the Individuals with Disabilities Education Act and its 2004 reauthorization.

A number of test-taking skills resources were available to college students. Some were devoted exclusively test-taking (Sides & Korchek, 1998; Hoefler, 1995). Others devoted a section or entire chapter to strategic test-taking (Beatrice, 1995; Dembo, 2004; Downing, 2005; Ellis, 2006; Ferrett, 1996; Ferrett, 2010; Kanar, 2004; Langan, 2001; Reynolds, 1996; Wong, 2006). Some specifically addressed the learning and assessment needs of nursing students (Alfaro-LeFevre, 1995; Chenovert, 2006; Hoefler, 1995; Katz, 2004; Sides & Korchek, 1998). For this review, a representative, rather than exhaustive, sample of sources was chosen (See Appendix A). The texts represented a 15-year publication range (1995 to 2010) in order to observe if the recommended approaches have changed. The collection was a convenience sample in that all were sources available to faculty, staff and students in a college of nursing. The sources were also chosen because they represented an extensive range of number of editions in print (one to seven). Some, therefore, were first-time publications for the authors, while others had sold multiple thousands of copies over numerous releases of new, updated editions, suggesting that instructors and students had found them to be beneficial over time and continued to adopt and purchase them.

Commonalities. The literature that addressed test-taking often included numerous kinds of test questions—essay, fill-in, short answer, in addition to multiple choice. The sources reviewed for this study addressed, either in whole or in part, strategies for taking multiple choice tests. Several commonalities appeared in the

academic support resources relative to multiple choice test-taking. Table 2 addresses factors that are reflected in the complete strategy of M4Q. Gray connotes nursing texts.

Table 2

Test-taking Advice in Selected Texts1995-2010: Factors in Common with M4Q

		Decide on the answer					
Pub. year	Book emphasis	…before reading the options	…after reading the options	Find key terms or key words	Rephrase or ask what it's about	Relate to Bloom's levels	Author
1995	Critical thinking	x		x			Beatrice
1995	Nursing			x			Alfaro-LeFevre.
1995	Nursing			x	x		Hoeffler
1996	Strategies		x				Reynolds
1996	Study skills						Ferrett
1998	Nursing		x				Sides & Korchek
2001	Skills		x				Langan
2004	Motivation, strategies.		x		x		Denbow
2004	Nursing		x	x			Katz
2004	Nursing	x					Rayfield & Manning
2004	Strategies	x	x				Kanar
2005	Nursing		x	x			Sylvestri
2005	Strategies	x					Downing
2006	Nursing		x				Chenovert
2006	Strategies	x	x				Ellis
2006	Study skills						Wong
2007	Nursing, alternative format tests			x		x	Sylvestri & Mojica
2010	Strategies	x			x		Ferrett

Table 3 repeats the texts and their dates, but addresses often-repeated advice that was not reflected in applying M4Q. The NCLEX is computer interactive, so some of the recommendations are not possible, and are so noted.

Table 3

Test-taking Advice in Selected Texts1995-2010: Contrasts to M4Q

			Not Possible on Computer Tests				
Pub. Year	Book Emphasis	Eliminate options one-by-one	Underline/ Mark	Circle	Cross out	Skip	Author
1995	Critical Thinking		x	x			Beatrice
1995	Nursing						Alfaro-LeFevre
1995	Nursing		x				Hoeffler
1996	Skills & Strategies	x					Reynolds
1996	Study Skills		x				Ferrett
1998	Nursing	x					Sides & Korchek
2001	Skills				x	x	Langan
2004	Motivation, Strategies.						Denbow, M. H.
2004	Nursing	x					Katz
2004	Nursing						Rayfield & Manning
2004	Strategies	x					Kanar
2005	Nursing	x					Sylvestri
2005	Strategies	x					Downing
2006	Nursing	x	x				Chenovert
2006	Strategies	x					Ellis
2006	Study Skills						Wong
2007	Nursing, Alternative Format Tests						Sylvestri & Mojica
2010	Strategies						Ferrett

Lack of evidence. Unlike the texts in use in virtually any other academic discipline**,** none of the numerous textbooks found that were written to teach students to become better scholars in general, and better test-takers specifically, contained citations with references that supported the advice or claims. Sylvestri and Mojica (2007) used references at the end of each chapter, but with no indication of whether the content of the advice or the exact practice test questions in the chapter were being referenced. In some instances—getting enough rest the night before a test, for example, one might reasonably assume that conventional wisdom is obviously correct. In other instances, the authors may assume that supportive evidence is so well known it does not need to be cited or repeated

Differences. The most striking difference between the approaches promoted in the test-taking literature is what the test-taker is encouraged to do next after initially reading the stem: read the options or not read them. This consideration was a key factor in developing M4Q. Reading the options before thinking about an answer gives each distractor the power to distract. Other differences may reflect the various authors' areas of expertise or interest. Without documentation, there is no way to determine the source.

Reading or not reading the options. The most frequent suggestion in both the older and more recent texts was to read all the options, then begin making decisions about which ones to eliminate (test each option by re-reading the stem, then immediately following it with a re-reading of the option, one at a time. Katz stated, "Your goal is to leave yourself with two possible answers, which would give you a 50-50 chance of making the right choice" (p. 253). As early as 1998, Sides and Korchek had also urged test-takers to reduce the options to two, so that is ongoing advice.

Four of the texts that were examined advised students *not* to read the options immediately after reading the stem. “On a multiple-choice exam, try to answer the question in your head before looking at any of the choices” (Beatrice, p. 237). “Read the question and decide on an answer first; then look for that answer in the list of choices” (Downing, 2005, p. 221). “Formulate your own answer before you read the answers given” (Chenovert, 2006, p. 91). “Cover the answers (called ‘distracters’) as you read the question and see what answer first comes to you. Then, look at the answers to see if your answer is one of the choices” (Ferrett, 2010, p. 270). Of the four, only Chenovert (2006) was specific to nursing education. The M4Q strategy also took that approach.

Purposeful advice. A few authors proposed strengthening the test-taker’s intentionality and purposefulness through a deeper awareness of the testing process. Wong (2006) stated that there are four sources of response to a test question. The first is Immediate: the examinee knows the answer. The second is Delayed: the examinee needs to re-read it and think about it before s/he can answer. The third is Assisted: the examinee needs to skim through everything and look for associations. The fourth is Educated Guessing, which she says should only be used when all else fails.

Kanar agreed with Wong’s intentional approach, and states three ways to take a multiple-choice test question:

1. If you know the material, first answer the question mentally and then read all the options and choose the correct one.
2. If you know the material but cannot answer the question mentally, read the options, eliminate those you know are incorrect, and choose the answer from those remaining.

3. If you do not know the material, or if you cannot figure out the answer, guess (2004, pp. 299-300).

"Unique" Advice. In several texts, unique and sometimes conflicting advice was given that had the appearance of being a trick, common knowledge, or individual opinion. "If you haven't a clue, choose "C" (Ferrett, 1996, p. 212). "Once you have read a question, look at the choices and try to answer the question. This strategy reduces the possibility that the choices will confuse you" (Katz, 2004, p. 253). "The longest answer is often correct" (Langan, 2001, p. 138). "The most complete and inclusive answer is often correct" […] "An answer in the middle, especially if it is longest, is often correct" (Langan, 2001, p. 138). None of these were supported with evidence.

Nursing compared to non-nursing. A few of the constructs showed distinct contrasts. The contrast between nursing-specific and non-nursing specific texts is evident in Table 1. Among the generic texts examined, six (35%) recommended using critical thinking to decide on the correct answer to the question before reading the distractors. Of the eight nursing-specific texts examined, only two recommended thinking about the answer before reading the distractors. Four of the 17 texts examined urged the test-taker to think conceptually or in terms of main ideas or principles, and they were all nursing-specific texts.

The other factor that showed a high level of contrast between generic and nursing-specific texts was the pattern of offering unique, often questionable, advice without evidence. One example is always choosing a particular letter to answer questions to which the answer was not known. Only two (25%) of the eight nursing-specific texts demonstrated that practice (Hoeffler, 1995; Katz, 2004), while five of the nine non-

nursing books, spanning the entire time range of the sample, did (Beatrice, 1995; Langan, 2001; Downing, 2005; Ferrett, 1996; Ferrett, 2010).

Age of advice. By viewing the resources were chronologically by year of publication some apparent trends emerged. The strategy of eliminating distractors one at a time is congruent with the strategy of deciding on an answer after reading all the options, and the pattern of their appearance in the texts is consistent. Only one of the six books that suggested deciding on an answer to the question before reading the options was written before 2004. The advice to decide on an answer after reading the options was common in both nursing-specific and non-nursing-specific texts from 1996 to 2006.

Academic folk wisdom. Skinner (2009) was challenged by the dearth of evidence for many of the commonly-held academic beliefs. Drawing from his own earlier research, he published the results of a number of incisive quantitative studies of what he called "academic folk wisdom: fact, fiction and falderal" (p. 46). He researched, then empirically tested the following commonly-held beliefs:

Easy questions at the beginning of an examination "get students off on the right foot."

Don't change your first-chosen answer because it's probably correct.

Print multiple-choice examinations on paper of different colours to discourage cheating.

When in doubt, choose "c"

"I hope I shall never be deterred from detecting what I think a cheat" (Samuel Johnson, letter to James Macpherson, January 20, 1775, p. 298)

Humour has no place on examinations.

Students sitting in front/middle seats achieve higher marks.

Highest grades are obtained in morning classes.

"I do much better on essays than multiple-choice questions."

Students with unusual names earn poor grades (Skinner, 2008, pp. 46-49)

Skinner carefully analyzed the few studies in the literature that appeared to support the above statements. He found weaknesses in their research designs and intentionally avoided those weaknesses in his own studies. In the case of the statement about changing answers, for example, he suspected that if the data were considered using gender as a factor, the conclusion would be different, and it was. He found "that female students make more than three times as many correct-to incorrect changes as do males" (p. 47). This insight is especially significant in college academic programs that tend to be dominated by women, such as education and nursing. None of Skinner's studies supported the commonly-held beliefs he examined, leading him to conclude that commonly-held beliefs about academic strategies ought to be studied further, rather than unquestioningly accepted.

A potential remedy. In 2008, Dodeen published what he described as "the first comprehensive scale developed to assess test-taking strategies used by university students" (p. 409). He explained that such a scale can help not only student achievement, but instrument validity, as well. Students who have maximized their cognitive assessments of their own test-taking should be able to represent what they really know, with less contamination by ineffectiveness in test-taking. Dodeen's sample sizes were 50, 828, 553 and 235. He isolated 74 strategies or skills from a comprehensive literature

review, and divided them into four categories: Before-test, Time management, During-test, and After-test strategies.

Dodeen (2008) conducted ambitious studies for content validity, construct validity and criterion-related validity, as well as stability and overall reliability of his scale. In some instances, he considered emergent psychometric weaknesses and addressed them, then re-tested with different samples. He compared outcomes to other previously-validated instruments to measure validity of the new tool. After exhaustive statistical analysis, he challenged others to "replicate and validate the scale using different samples from different educational levels" (p. 418). A scale that could accurately assess the strategies students use, with validity and reliability, could be a significant step toward bringing strategic test-taking out of the realm of academic folklore and into the realm of academic science.

Conclusion and Recommendation

The review of the literature supports the belief that there is a great need for well-designed and rigorously conducted research in strategic test-taking related to the theoretical considerations as well as the quantitative outcomes. The psychometric values should be set at standards as demanding as other academic research. Tools should be validated. The research should be ongoing, and the results should be disseminated, especially considering the prevalence of high-stakes testing at all levels of education. In an academic major such as nursing, in which entry into the profession hinges on passing a comprehensive, high-stakes multiple-choice examination after graduating from college with a nursing major, the literature supports a particular need for evidence-based test-taking strategies.

CHAPTER 3: METHODOLOGY

The purpose of Chapter 3 is to explicate the components of the methodology for this study. The chapter is divided into the following sections: Research Design, Sampling Procedures, Instrumentation, Intervention, Data Collection Procedures, Data Analysis Procedures, and Threats to Internal Validity. The model for the Mayfield's Four Questions test-taking strategy was included under Intervention, and how the subjects were protected from harm was addressed under Sampling Procedures.

Research Design

The purpose of this study was to determine whether or not there is a statistical relationship and/or correlation between one independent variable: learning the Mayfield's Four Questions (M4Q) multiple-choice test-taking strategy, and two dependent variables: (1) nursing students' perceived self-efficacy, and (2) their grades on multiple-choice tests in their nursing courses. The research design was characterized as a Randomized Subjects, Pretest-Posttest Control Group Design (Ary, Jacobs, Razavieh, & Sorensen, 2006). The quantitative data consisted of results from pre- and post-intervention administrations of the General Self-Efficacy Scale (GSE) (Schwarzer & Jerusalem, 1993) and the participants' grades on all multiple-choice tests administered in their nursing courses during the study. Experimental and control groups were designated by standard random sampling methods with strata delineators, which were fixed constants. The statistical analysis addressed individual, class, and research group correlations and differences in subjects' responses to the 12 questions on the perceived self-efficacy survey individually and by group. It also addressed the individual and

aggregate quantitative data from in-class multiple-choice test grades relative to receiving or not receiving the intervention of being taught M4Q.

The research proposal was submitted to the Institutional Review Board (IRB) of the degree-granting university and approved (see Appendix A). The proposal was also submitted to the IRB of the college at which the researcher is employed and approved (see Appendix B). The research site college did not require additional IRB approval.

Sampling Procedures

Site

The research site was a private, denominational college located in a small city in the Midwest. The college was established in 1858, and the nursing program was added in 1985. At the beginning of the academic year 2009-10, the enrollment at the college was approximately 1,150 students. The enrollment in the nursing program in the fall semester was 54. Based on the previous semester's enrollment, the attrition and persistence history of the program, and the fact that the cohort of new Licensed Practical Nurse associate degree students must always begin their curriculum in the spring semester, the expected enrollment for the semester of the study was approximately 70.

Population

The population for the study was all the associate degree nursing majors, therefore the sample consisted of homogeneous groups of first- and second-year nursing students. The subjects' official level in the nursing program was a fixed constant, and was designated in the study by the terms used in the institution—*first-year* and *second-year*, rather than *freshman* and *sophomore* as proposed.

Sampling

Stratified random sampling was used to select an experimental group and a control group for the study, with *first year* and *second year* used as the strata delineators. The director of the nursing program invited all students in the population to meet together at a high-availability time and convenient place on campus to be offered the opportunity to participate in the study. At the meeting, the researcher introduced the study. All students were given consent forms (see Appendix C) and the General Self-Efficacy Scale (GSE) (Jerusalem & Schwarzer, 1993) (see Appendix D) that had been appropriately adapted as a pre-intervention survey. Students submitted the hard copies that day and the next, and the 51 students who turned them in became the participants in the study.

The submitted signed hard copies were sorted according to strata delineators then randomly selected to the control and experimental groups. That selection method ensured that each group had similar representation of first- year and second-year students. Participants were notified whether they had been randomly selected for the control or experimental group by email and by notes in their college mailboxes. Subjects in the experimental group were asked to contact the researcher to set up individual appointments to begin the intervention. Numerous available times were offered, and telephone numbers and email addresses that could be used to contact the researcher were given. Immediately after the notifications were sent, one subject in the experimental group withdrew from the study. For the beginning of the intervention, therefore, the control group consisted of 14 1st year students and 11 2nd year students. The experimental group consisted of 16 1st year students and 10 2nd year students.

Several factors contributed to a significant decrease in the number of subjects whose data was subjected to the final statistical analysis. In addition to the one subject who withdrew from the study immediately after signing the consent form, three others formally withdrew after some level of participation in the experimental group. Five subjects randomly chosen for the experimental group signed the consent form and completed the initial pre-intervention survey, but did not participate in the intervention, and were therefore dropped from the final data. One subject was dismissed from the college during the study, and another withdrew. Others either did not complete one of the surveys or missed more than one course test that was not made up and were dropped from the final data considered for analysis.

The final groups consisted of 16 subjects in the experimental group and 17 in the control group who had taken both the pre-intervention and post-intervention surveys and missed no more than one in-class test. In addition, to satisfy the operational definitions for the study, each subject in the final experimental group must have attended at least two meetings with the researcher for the purpose of learning to apply the M4Q strategy to multiple-choice questions.

Subject protection. The subjects were protected from harm in that they received full disclosure of the potential risks orally in the group setting in which they were invited to ask questions, and in writing on their consent forms (see Appendix C). They were further protected in that and their personally-identifiable data was kept confidential, and the statistical analysis was conducted using blind data. There was no risk of harm to the control group because they had every opportunity to be successful on their tests that they had before the study.

Pre-emptive protection. The Experimental Group was at risk in that the intervention could have resulted in lower test scores than the subjects would have earned without the intervention. The potential for that risk was identified from anecdotal reports from nursing students in another college who had used the Mayfield's Four Questions (M4Q) test-taking strategy, the core of the intervention. The students in the other college had applied the strategy in a testing situation before they had practiced it enough to gain proficiency, then they ran out of time before they could complete the test. That risk was controlled in this study by teaching students about that risk and suggesting appropriate ways to gradually increase use of M4Q as proficiency increased, rather than taking an all-or-nothing approach to its application in real testing settings.

Institutional Review Board (IRB). The research proposal for this study was submitted for IRB approval from the degree-granting university (see Appendix A). IRB approval was also received from the IRB affiliated with the researcher's employer, Blessing-Rieman College of Nursing (see Appendix B) . Since the Blessing-Rieman IRB is responsible for research conducted in a regional medical center that receives government funding as well as the college in which the researcher teaches, the review was different from the university's and rigorous in different ways. Approval from two such IRBs further supported the subjects' protection from harm.

Data Collection and Protection

The consent forms and pre- and post-intervention surveys were administered by the researcher, collected and secured. The nursing department's secretary at the research site college was provided with copies of all consent forms to archive. Throughout the semester, all subjects' test grades were sent by the nursing faculty members directly to the nursing department's secretary. The researcher provided her with a dedicated

portable data saving device for research data that was returned to the researcher after the final data had been entered.

Data was gathered for the percentage grades earned on all formative and summative course tests and institutionally-required examinations administered during the semester, shown in Table 4. The instruments were listed by the name used by the site program and recorded by the nursing department secretary in the raw data. HESI 1 for first-year students was Fundamentals, while HESI 1 for second-year students was Pediatrics. Means, standard deviations, and standard error of the means were calculated for each of the questions by group and by level in the groups.

Table 4

Multiple-Choice Tests Data Sources

Fixed variable	Type of test	Curriculum content/title
First-year students:	Formative:	Obstetrics 1, 2, 3 and 4
		Test 1, 2, 3, and 4
		Med/Surg 1, 2 and 3
		Nutrition 1 and 2
		Demo 1, 2, 3, 4 and 5
	Summative:	Final Exam
		HESI Fu
		HESI M
Second-year students:	Formative:	Test 1, 2, 3, 4, 5, 6, 7 and 8
		Nutrition 1 and 2
		Pharmacology 3 and 4
	Summative:	Final Exam

HESI Pe
HESI Ac
Exit Exam 1
Exit Exam 2

The proposal for this study stated that other details would be determined in collaboration with the director of the nursing program at the time of the study. That was done. Semester grades were not retrieved from the Office of the Registrar, as indicated in the proposal for the study because the director preferred to work within the department and offered the time and services of the secretary to facilitate accurate and timely data collection. The nursing faculty gave a significant number of multiple-choice tests at both levels, and those test grades constituted such a large proportion of the students' semester grades, the additional data of the final course grade would have been extraneous and redundant. It was, therefore, not statistically considered.

The researcher created a data base for the survey data in Excel spreadsheets which were stored on a portable data storage device and backed up on hard drives in two locations, both password protected. Electronic data spreadsheets were kept with student names, so that students who chose to withdraw from the college or from the research project could have all their data removed without affecting the other data. The secretary assigned random numerical identifiers to all subjects and also retained a second data set of spreadsheets with only those identifiers. After the research was completed and the database was finalized, the data set without names was used for the statistical analysis, increasing confidentiality of the participants.

All subjects who were missing either the pre- or post-survey or more than one formative test were removed from the cohorts. Based on the working definition of "were

taught the M4Q multiple-choice test-taking strategy" for this study, all experimental groups subjects who did not meet with the researcher at least twice to learn M4Q were also removed from the cohorts. The final cohorts for statistical analysis were experimental group (n=16) and control group (n=17).

Instrumentation

The General Self-Efficacy Scale (GSE) (Jerusalem & Schwarzer, 1993) was adapted for use for the pre- and post-intervention surveys in this study. The authors of the GSE stated that "in most applications it is necessary to add a few items to cover the particular content of the survey or intervention" (Schwarzer, 2008, p. 1). Bandura (2006), who developed Self-Efficacy Theory, also recommended increasing the specificity of general self-efficacy scales and published advice on how to improve them. (See Chapter 2.) Consistent with the expert advice, two questions were added which directly addressed the variables being tested in the study.

The GSE was written and published in German about 1980: sources variously list the date as 1979 and 1981 (Schwarzer, 2008; RCMAR Measurement Tools, 2006), and later publication dates appear for various translations. The authors are variously cited in reverse order, as well. The English translation was published in 1993 (Schwarzer, 2008) The GSE has been subjected to validation studies in at least 25 countries in a wide range of constructs and is available online for researchers' use in 27 languages (RCMAR Measurement Tools, 2006; Schwarzer, 2008). Following a comparative validation study of three self-efficacy instruments including GSE, Scherbaum, Cohen-Charash and Kern (2006) concluded that "the items on these measures function fairly well and demonstrate many desirable psychometric properties" (p. 1058). Based on the supporting research

and recommendations, the adapted GSE was considered adequate and appropriate for this study.

The pre- and post-intervention surveys were identical except the prefixes pre- and post- in the title, and the pre-intervention survey was presented as Appendix D. The survey was set up as a five-point Likert scale, and consisted of 12 statements, the first 10 original to the GSE. The survey statements were:

1. I can always manage to solve difficult problems if I try hard enough.
2. If someone opposes me, I can find the means and ways to get what I want.
3. It is easy for me to stick to my aims and accomplish my goals.
4. I am confident that I could deal efficiently with unexpected events.
5. Thanks to my resourcefulness, I know how to handle unforeseen situations.
6. I can solve most problems if I invest the necessary effort.
7. I can remain calm when facing difficulties because I can rely on my coping abilities.
8. When I am confronted with a problem, I can usually find several solutions.
9. If I am in trouble, I can usually think of a solution.
10. I can usually handle whatever comes my way (Jerusalem & Schwarzer , 1993).

The two additional questions added to provide context were:

11. I consider myself to be proficient at answering multiple-choice test questions.
12. In addition to content knowledge, I have strategic skills as a test-taker.

Intervention

The intervention consisted of providing individualized teaching and guided practice of the Mayfield's Four Questions multiple-choice test-taking strategy to the

subjects in the experimental group. The intervention occurred in weekly 30-minute appointments. Students randomly selected for the experimental group were invited to set up their own appointments via email. A wide range of morning, afternoon and evening times were offered on Mondays through Thursdays, February through April, at the site college and in a city about 20 miles away in which the researcher had an office.

Content

The content of the intervention was a specific, formalized test-taking strategy, Mayfield's Four Questions (M4Q), which had not previously been validated by experimental research (see Appendix E). M4Q consisted of a sequence of four questions the test-taker asked himself/herself about a multiple-choice question under consideration, before reading the options:

(1) What is this question really asking me?

(2) What is it trying to determine that I know?

(3) What level of learning is required to answer correctly?

(4) What do I think is the correct answer?

In the first session, each subject received a handout that explained M4Q and received a detailed verbal explanation of the strategy (See Appendix E). Then the subjects began practicing its application, with directive coaching by the researcher, using a book of sample nursing-specific multiple-choice test questions. The researcher explained that the purpose of subsequent meetings would be ongoing practice for mastery in applying the strategy.

The books of test questions the researcher provided for use in the study were six copies of *Test-taking Skills for Beginning Nursing Students* (Nugent & Vitale, 2005). The Nugent & Vitale text had been used as a learning tool for teaching this strategy for

many years prior to this study, but in another college. It was carefully chosen from among many options because of its user-friendly format, the organization of the content, the nursing level of the content, and the fact that rationales were given for the incorrect options as well as the correct ones, so learning could take place from both kinds of rationales. Two subjects used books of their own choosing for at least one appointment each.

Conceptual Design

Test-takers who applied the M4Q strategy thought about the question in the ways directed by the four questions as represented in the Conceptual Design. Moving through the four questions took the reader through all five levels of thinking represented in Bloom's Taxonomy of Learning Objectives (Bloom, 1956). Identifying important words and recognizing the intent of the question tested the students' basic knowledge and comprehension of the topic at hand.

The first step in using M4Q was to carefully read the multiple-choice question and decide what was being asked. In assessment nomenclature, the entire question with the possible answers is called the *item*. The part of the item that requires the test-taker to make a decision or answer a question is called the *stem*. The possible answers are called options, and the correct answer(s) is/are called the *correct answer(s)* or *correct response(s)*. The incorrect answer(s) is/are called *distractors* (sometimes spelled *distracters*).

Next, the test-taker asked, "What is it trying to determine that I know?" Critically thinking about this question helped the student place the content of the question into the context of his/her knowledge. Students often thought what the question asked is also

what it was trying to determine was known, but that is most often not the case. A scenario question might actually be assessing the nursing student's knowledge about Maslow's Hierarchy of Human Needs, Erickson's Developmental Stages, protocols for sterile techniques, priorities for treatment in an emergency, or the stages of grief. Answering the second M4Q question clarified that for the students.

The third question was therapeutic and pre-emptive. Observations over many years supported the premise that students who knew more than was required to answer a question correctly often read into it and/or perceived it as more complex than it was. Analyzing the question relative to Bloom's Taxonomy reassured the test-taker that s/he needed only to think at the level of Application or lower for a multiple-choice question in nursing. That observation in no way minimized the importance or difficulty of the questions, but the higher levels of thinking—synthesis and evaluation, must be assessed via different tools, i.e. research and writing prose.

Finally, the student was asked to determine what s/he thought was the correct answer, and look for it among the options. After identifying it, the student was encouraged to skim through the other options to make sure there was not an umbrella statement that was a better choice. An umbrella statement is true and is broad enough to include or cover other options offered. If there was not one present, the student was strongly encouraged to have confidence in the chosen response and select it as the answer. Since the distractors were never seriously studied and considered, they lost most of their power to distract. The M4Q strategy facilitated the discrimination between those who did and those who did not know the correct answer to the questions, therefore strengthening the validity of the students' tests.

The Conceptual Design for Mayfield's Four Questions shows the critical thinking sequence involved in applying the strategy. Each of the four questions challenges the test-taker at an increasingly higher level of Bloom's Taxonomy (1956) and logically leads the student to decide on an answer synthesized and retrieved from his or her knowledge base.

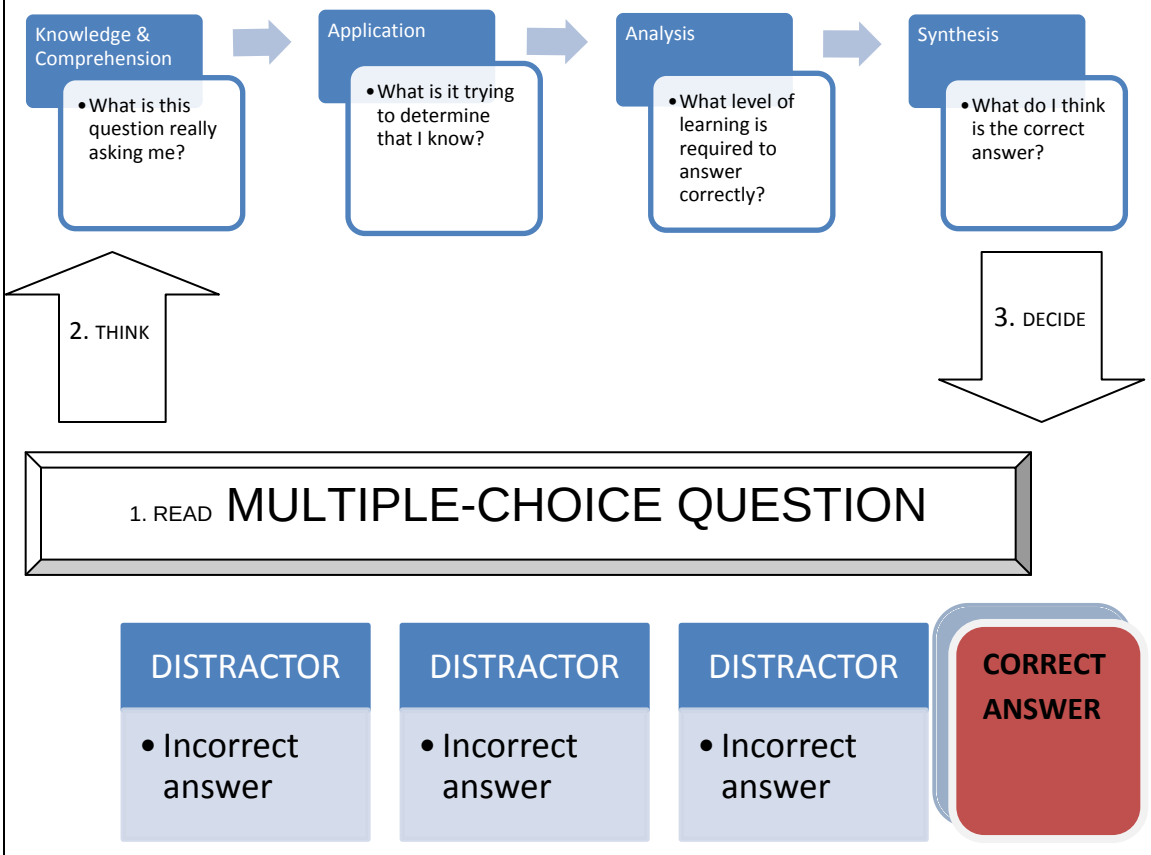

Figure 1. Conceptual Design: Mayfield's Four Questions (M4Q)

An important caveat in the M4Q test-taking strategy was that the test-taker should not look at the options until the four M4Q questions have been answered. At each meeting, therefore, the students were offered their choice from a variety of small cards and papers to use as a question cover. In the meetings the first week after mid-term break, the researcher gave each experimental group subject a small laminated card with the M4Q questions on one side and Bloom's Taxonomy and an umbrella image on the other. From that point on, subjects decided whether to use the instruction handout, a blank card or paper, or the laminated M4Q cards, as their M4Q learning props. All chose to use the laminated cards.

The researcher used a form to take notes for each meeting with a subject, but beginning the first week of appointments, students came in groups instead of individually. The researcher immediately recognized that recording the students' names, the date, level, and any anecdotal notes from the students' comments would take too much of the allotted meeting time. The researcher adapted the standard intervention procedure to that observation, and from then on handed the form to the students upon arrival and asked them to complete the data required and make any notations they wished to share. The researcher often added notes after the meeting, particularly noting the beginning and ending questions that had been practiced with M4Q during the session, which expedited getting started promptly and in the correct place in the questions book at the next meeting.

Threats to Internal Validity

Fraenkel and Wallen (2003) identified types of threats to internal validity related to subject characteristics, location, subject mortality, and instrumentation. Those threats

and others were adequately addressed by the design and implementation of this study. The subject characteristics threat was a minimal threat and was addressed through random sampling from the two academic levels in the nursing program. The location threat was considered to be potentially strong, but was not manifested and therefore did not need to be addressed.

Subject mortality. The subject mortality threat, which is common in intervention studies, was significantly realized and impacted the outcome of the study. That threat could not be directly controlled. Seven students who agreed to participate, signed a consent form, and were randomly assigned to the experimental group never made an appointment. Three of the seven formally withdrew from the study, but the other four did not. Three other experimental group participants only kept two appointments. One student in the control group was dismissed from the college and one withdrew.

History. History threats were minimized by communication and cooperation with the personnel of the research site. The faculty, staff and students were positive about the study, with the exception of a few students who strongly wanted to be in the experimental group and receive the intervention, but were randomly selected for the control group. Maturation was not a threat because the intervention was only for one college semester, and all subjects were already adults.

Instrumentation. Instrumentation offers several types of potential threats. Since the pre- and post-intervention survey instrument was an adaptation, no validity or reliability had been established for the final two questions. The first 10 questions, however, had been extensively tested and validated.

The data collector bias threat was potentially the greatest threat, since the researcher personally delivered the intervention, and the subjects' professors who taught their classes and developed, administered and graded the tests whose scores formed the data, were also data collectors. Since the test data were all from multiple-choice questions, and the students were in the same classes with each other, there was no threat of decay due to subjective grading.

Standardized procedures for data collecting, recording and analyzing addressed the data collector bias threat. That threat was also addressed by not allowing the instructors to know which students were in the control group and which are in the experimental group until it became apparent to them by observing their behaviors in testing situations. The data set without names, but only randomly-assigned identifiers assigned by the secretary, was used for statistical analysis.

In addition to the threats documented by Fraenkel and Wallen (2003) the threat of the Hawthorne effect (Massie, 1971) existed. When the Hawthorne effect is manifested, subjects respond positively due to the extra attention they received as research participants, regardless of the intervention. When testing a new strategy in a small educational setting, the Hawthorne effect may be unavoidable. Future study in a larger population with a design in which there are multiple individuals performing the intervention is warranted.

Data Analysis

A database for the study was created using Microsoft Excel 2007 spreadsheets. Spreadsheet files were maintained by group for the pre-and post-intervention survey responses and for all multiple-choice tests administered to the subjects in their nursing

classes during the study. The Statistical Package for the Social Sciences 18 (SPSS) (temporarily and interchangeably referred to by copyright owner as PASW 18) software and functions of Microsoft Excel 2007 were utilized for the statistical analyses of the data. A significance criterion of .05 for the confidence level was specified. Statistical analysis included the one-sample *t*-test, which compares the means of two variables and reports the degree of statistical significance; the paired t-test, which allowed comparison of the groups' pre- and post-intervention responses; and the analysis of variance (ANOVA), which compares the means of more than two variables and reports the degree of statistical significance. The SPSS execution of the *t*-test also produces correlation coefficients and reports their degree of statistical significance.

Comparison of Proposed and Actual Data Analysis Designs

The Ruban, McCoach, McGuire and Reis (2003) design for statistical analysis was chosen as the model for the M4Q data analysis as it was proposed because it contained several similarities with the M4Q study: (1) The theoretical framework for both studies was social cognitive theory with a focus on self-efficacy. (2) The authors of the Ruban et al study stated that their study "provided compelling evidence that a strong link exists between students' academic self-regulation—particularly their use of learning strategies—and their academic achievement" (p. 271). The M4Q study addressed a specific learning strategy-- test-taking skills, and examined the effect learning that strategy had on academic achievement. (3) Both studies tested the relationships between specific academic skills and grades of college students. (4) Both compared the data with fixed strata indicators representing the level in the academic program. (5) Both models utilized a 5-point Likert scale survey instruments.

The Ruban et al study utilized a new assessment tool, the *Learning Strategies and Study Skills* survey (LSSS), which was developed by the researchers "to assess college students' self-reported study behaviors in generic learning situations" (Ruban et al, 2005, p. 275). The M4Q study utilized the widely-validated General Self-Efficacy Scale (2008) to assess students' perceived self-efficacy before and after an intervention. With content-specific modifications, the statistical model designed by Ruban, McCoach, McGuire and Reis (2003) was considered adequate to support or reject the null hypotheses of the M4Q study.

Descriptive treatment. The Ruban et al design first applied descriptive statistics to demographic variables present in the population. Ruban et al specified five variables: Gender, Age, Ethnicity, Academic level, and GPA, which were all compared to the variables of "With learning disabilities" and "Without learning disabilities." The M4Q study specified four of the same of the first five variables, excepting only Age, and replaced the Ruban et al fixed variables of "With learning disabilities" and "Without learning disabilities" with the independent variables of "With intervention" and "Without intervention" for comparison.

Survey responses. The adapted General Self-Efficacy Scale (GSE) (Jerusalem & Schwarzer, 2008), was the instrument utilized for the pre- and post-intervention survey. It was composed of ten widely-studied and validated questions and two additional questions to make it more specific to the construct of multiple-choice test-taking. The only difference between the pre- and post- surveys was the heading that designated it as such. All subjects who (1) completed both the pre- and post-intervention surveys, (2) completed, with not more than one exception, all the multiple-choice tests given in the

nursing courses for the subject's level, a final total of 16 in the experimental group and 17 in the control group. All scores on all survey statements were documented and the SPSS analyses generated the descriptive statistics. They were not, however, necessary for testing the two hypotheses.

Test scores. Descriptive statistical analysis was conducted to analyze the subjects' multiple-choice test scores. Resulting data included the means that were the foundational statistics for testing the hypotheses.

Omnibus Run and Multiple Group Analysis. Ruban et al designed their study to address the relationships between learning strategies, conceptual skills, memorization skills, compensation strategies and the subjects' grades as expressed by a grade point average (GPA). They found that

> because the sample of students with LD was so small, the analyses possessed inadequate power to statistically test the plausibility of imposing equality constraints across groups. However, a visual inspection of the standardized structural parameters and R^2 for the two groups revealed some interesting differences between the groups (p. 277).

Likewise, the small sample size (n=33) in the M4Q study necessitated visual inspection for some comparisons and a different statistical analysis than Ruban et al used for others.

Factor analysis. The third part of the Ruban et al design was a factor analysis that examined the relationships between 19 dimension/item stem factors and (1) conceptual skills, (2) routine memorization and (3) compensatory supports. The purpose of that analysis was to obtain support for the construct validity of the new instrument. Since the instrument utilized for the M4Q study had been validated by more than 25

studies in 27 languages, and further validation was not a purpose of this study, that factor analysis was not attempted. To demonstrate the logical flow of activity in the intervention, a conceptual model was developed, rather than a factor analysis. (See Figure 1)

Comparisons and correlations among the variables. The fourth part of the Ruban et al design was determining comparisons and correlations among the variables. Due to the great number of variables in that study, more complex correlational statistics were applied than were required to calculate robust correlations for the M4Q study. The comparisons between the survey variables in the M4Q study were explored using various *t*-tests appropriate for comparing two means, and the analysis of variance (ANOVA) when means of more than two factors were compared. Correlations between the means of the variables expressed as Pearson *r* values are unspecified secondary data outcomes calculated by SPSS and reported as part of the *t*-test reports. The variables utilized as statistical factors in the study were group, level, test scores, pre-intervention survey scores and post-intervention survey scores. The *t*-tests and ANOVAs examined the group differences and their significances as reported in *t*-scores; the correlations examined the relationships and their significances. Field (2006) noted that the effect size is measured with the Pearson's correlation coefficient calculated in the SPSS *t*-test, and suggested the following paramenters:

r = .10 (small effect), in this case the effect explains 1% of the total variance

r = .30 (medium effect), the effect accounts for 9% of the total variance

r = .50 (large effect), the effect accounts for 25% of the total variance (2006, p. 32).

Structural modeling. The fifth part of the Ruban et al design was structural modeling between the factors that had been statistically compared. Only five of the total resulted in statistically significant paths: perceived usefulness, conceptual skills, routine memorization, GPA and compensation strategies. The others were dropped from the Ruban et al model. Chi-square is a statistical model used to test frequency data, and while similar to the *t*-tests for testing the difference between sample means, the approaches are not interchangeable (Welkowitz, Ewen & Cohen, 2000). The appropriate *t*-test statistical applications for testing means of samples, rather than observed frequencies, were therefore appropriately applied in the M4Q study. A path analysis structural model was not appropriate for the M4Q study.

Summary of Data Analysis Design

The first group of data presented by Ruban et al was descriptive and comparative between the experimental and control groups. The methods were visual inspection and means of groups. Those methods were employed in the M4Q study. The second through fifth components of the Ruban et al design employed visual inspection and calculations of statistical comparisons between multiple factors. In the Ruban et al study, the numerous factors necessitating using factor analysis and step-wise regression. Because of the homogeneous nature of the final data for the M4Q study, neither factor analysis nor step-wise regression were appropriate statistical applications. The statistical analyses utilized for the M4Q study were limited to the reported results of the *t*-tests and ANOVAs and the correlation statistics that were reported as data extraneous to, but supporting, those tests.

CHAPTER 4: RESULTS

Two null hypotheses were tested in this study relative to the intervention of teaching nursing students the Mayfield's Four Questions (M4Q) multiple-choice test-taking strategy. One hypothesis concerned the subjects' self-efficacy with and without and before and after a multiple-choice test-taking strategy intervention. The other concerned their performance on multiple-choice tests before and after the test-taking intervention. The self-efficacy dependent variables were measured by administration of five-point Likert scale pre- and post-intervention surveys to all subjects in the experimental and control groups. The results of the two surveys were analyzed by visual inspection and were analyzed by group, strata, and subject using the data generated by running appropriate statistical analyses in SPSS.

The data for the test performance variable was expressed in means of percentage-of-correct scores on all formative multiple-choice tests in nursing courses and all summative multiple-choice nursing tests administered at the end of the semester. The data included scores on instructor-written formative and summative tests and standardized summative assessments. The test grade data was subjected to analyses by visual inspection and SPSS *t*-tests and ANOVAs. The means of groups, strata and subject variables were compared.

The Ruban, McCoach, McGuire & Reis (2003) research design selected as a model for the M4Q study was multi-faceted, and was adapted for this study. This section delineates the results of the adapted analysis.

Descriptive Statistics

Table 5 represents the M4Q study's adapted descriptive demographic data: gender, ethnicity, and academic level. The grade point averages (GPA) of the subjects of the M4Q study were not relevant to testing the hypotheses because GPAs are calculated including grades in all course work. The various components of the M4Q study's subjects' grades in courses other than nursing courses are unknown, and almost certainly contain assessments completed in methods other than multiple-choice tests, i.e. laboratory courses' skills grades. The relevant grade data for this study were limited to the multiple-choice test grades in the nursing courses. The means of those grades are shown with the demographic data in Table 5.

The final data in the M4Q study was too homogeneous to calculate statistics based on gender or ethnicity. Only one male subject was randomly selected to the experimental group, and only one subject of any ethnicity other than *White, non-hispanic* was randomly selected to the experimental group. Analyzing separate data for those two subjects would have unacceptably compromised their confidentiality. The results of the visual examination for demographics is shown in Table 5.

Table 5
Demographics for Experimental and Control Groups

Variable	Randomly Assigned Group		
	Experimental		Control
Gender	15	Female	14
	1	Male	3
Ethnicity	0	African-American	0
	1	Hispanic	0
	15	White, non-hispanic	17
Academic Level	7	1st Year	12

	9	2nd Year	5
Multiple-Choice Test Grades Group Means			
Formative Tests	81	Course Tests	82
Summative Tests	80	Final Exam	80
	77	HESI 1	74
	88	HESI 2	76

Null Hypothesis #1: Self-Efficacy Scale results. The data for the General Self-Efficacy Scale (GSE) statistical analysis consisted of each subject's response to each questions on the pre- and post-intervention survey. The questions on the pre-and post-intervention surveys were identical and were adapted from the English translation of the General Self-Efficacy Scale (GSE) (Jerusalem & Schwarzer, 1993) according to expert guidelines. The first ten statements comprised the widely-validated GSE scale, the last two were added in compliance with recommendations by the scale authors and self-efficacy theorist Albert Bandura (Bandura, 2002). The format was a 5-point Likert scale, ranging from Not at all true (1) to Completely true (5), with no intermediate descriptors. The questions were:

1. I can always manage to solve difficult problems if I try hard enough.
2. If someone opposes me, I can find the means and ways to get what I want.
3. It is easy for me to stick to my aims and accomplish my goals.
4. I am confident that I could deal efficiently with unexpected events.
5. Thanks to my resourcefulness, I know how to handle unforeseen situations.
6. I can solve most problems if I invest the necessary effort.
7. I can remain calm when facing difficulties because I can rely on my coping abilities.
8. When I am confronted with a problem, I can usually find several solutions.

9. If I am in trouble, I can usually think of a solution.

10. I can usually handle whatever comes my way (Jerusalem & Schwarzer (1993).

The two additional questions added to provide context were:

11. I consider myself to be proficient at answering multiple-choice test questions.

12. In addition to content knowledge, I have strategic skills as a test-taker.

For this study and the discussion of the statistical analyses, the statements in the pre-intervention survey were identified sequentially as Q1 through Q12. The identical statements in the post-intervention survey were identified in the same sequence as Q1X through Q12X, therefore Q1 was an identical survey question to Q1X, Q2 to Q2X, etc.

Visual inspection descriptive results. On the pre-intervention GSE survey, subjects in the control group reported higher mean scores than the experimental group on seven of the 12 statements. On the post-intervention survey, the experimental group reported higher mean scores than the control group on 11 of the 12 statements.

Statistical analysis: T-tests results. A one-sample *t*-test was calculated to compare the means of the experimental and control groups' individual statement scores on the pre- and post-intervention surveys to see if a statistically significant difference existed. The one sample was the group because each group retained its separate identity throughout the study.

Changes in experimental group's scores. The difference between the means of the pre-intervention and post-intervention individual statements on the experimental group's surveys was significant ($t(15) = 39.56$, $p = .000$), indicting a statistically significant change took place between the pre- and post-surveys (See Table 6).

Table 6

Experimental Group Pre- and Post- Self-Efficacy Survey Question Means

One-Sample Statistics

	N	Mean	Std. deviation	Std. error mean
MeansQ1to12 (Pre)	16	3.3813	.60797	.15199
MeansQ1Xto12X (Post)	6	3.9375	.39812	.09953

One-Sample Test

	Test Value = 0					
					95% Confidence interval of the difference	
	t	df	Sig. (2-tailed)	Mean difference	Lower	Upper
MeansQ1to12	22.246	15	.000	3.38125	3.0573	3.7052
MeansQ1Xto12X	39.561	15	.000	3.93750	3.7254	4.1496

Changes in the control group's scores. A one-sample *t*-test was calculated to compare the means of the control group's individual statement scores on the pre- and post-intervention surveys. The difference between the responses on individual statements was also statistically significant ($t(16) = 22.39$, $p = .000$) (See Table 7), but much smaller.

Table 7

Control Group Pre- and Post- Self-Efficacy Survey Question Means

One-Sample Statistics

	N	Mean	Std. deviation	Std. error mean
MeansQ1to12	17	3.4412	.45423	.11017
MeansQ1Xto12X	17	3.5941	.66188	.16053

One-Sample Test

	Test value = 0					
					95% Confidence interval of the difference	
	t	df	Sig. (2-tailed)	Mean difference	Lower	Upper
MeansQ1to12	31.236	16	.000	3.44118	3.2076	3.6747
MeansQ1Xto12X	22.389	16	.000	3.59412	3.2538	3.9344

Comparisons between groups. Visual inspection identified differences between the groups. On the post-intervention survey, the experimental group subjects indicated a self-efficacy score of at least 4.00 on six of the 12 statements. On the post-intervention (for them, end-of-study) survey, the control group subjects indicated self-efficacy scores below 4.00 on mean scores of all 12 statements. Field (2005) recommended utilizing the statistical technique called analysis of variance (ANOVA) when statistically comparing

the differences between more than two variables, rather than the *t*-test, so in addressing the survey statement individually, that recommendation was followed. A one-way ANOVA indicated no statistically significant difference between the group means on any of the 12 survey statements in the pre-test; but in the post-test, the experimental group scored higher than the control group with a difference that was statistically significant on three, or 25%, of the questions. The control group did not score higher than the experimental group at a statistically significant rate on any statement. (See Table 8.)

Table 8

ANOVA for Between-Groups Comparison of Pre- and Post-Intervention Surveys

GSE Survey Statement	Results of between-group statistical comparisons of pre-intervention and post-intervention survey statements
Q & QX 1 I can always manage to solve difficult problems if I try hard enough.	Pre: ($F(1,31) = 3.455$, $p = .073$) No significant difference between groups Post : ($F(1,31) = 1.600$, $p = .215$) No significant difference between groups
Q & QX 2 If someone opposes me, I can find the means and ways to get what I want.	Pre: ($F(1,31) = .03$, $p = .863$) No significant difference between groups Post: ($F(1,31) = .385$, $p = .540$ No significant difference between groups
Q & QX 3 It is easy for me to stick to my aims and accomplish my goals.	Pre: ($F(1,31) = .80$, $p = .378$) No significant difference between groups Post: ($F(1,31) = 1.941$, $p = .173$) No significant difference between groups
Q & QX 4 I am confident that I could deal efficiently with unexpected events.	Pre: $F(1,31) = .560$, $p = .460$) No significant difference between groups Post: ($F(1,31) = 11.809$, $p = .002$) **Significant difference between groups**

Q & QX 5 Thanks to my resourcefulness, I know how to handle unforeseen situations.	Pre: ($F(1,31) = .487$, $p = .491$) No significant difference between groups Post: ($F(1,31) = 2.679$, $p = .112$) No significant difference between groups
Q & QX 6 I can solve most problems if I invest the necessary effort.	Pre: ($F(1,31) = 1.468$, $p = .235$) No significant difference between groups Post: ($F(1,31) = 2.151$, $p = .153$) No significant difference between groups
Q & QX 7 I can remain calm when facing difficulties because I can rely on my coping abilities.	Pre: $F(1,31) = 1.058$, $p = .132$) No significant difference between groups Post: ($F(1,31) = 4.149$, $p = .050$) **Significant difference between groups**
Q8 & QX 8 When I am confronted with a problem, I can usually find several solutions.	Pre: ($F(1,31) = 1.05$, $p = .314$) No significant difference between groups Post: ($F(1,31) = .424$, $p = .520$) No significant difference between groups
Q & QX 9 If I am in trouble, I can usually think of a solution.	Pre: ($F(1,31) = .946$, $p = .338$) No significant difference between groups Post: ($F(1,31) = 1.417$, $p = .243$) No significant difference between groups
Q & QX 10 I can usually handle whatever comes my way.	Pre: ($F(1,31) = .829$, $p = .370$) No significant difference between groups Post: ($F(1,31) = 4.30$, $p = .047$) **Significant difference between groups**
Q & QX 11 I consider myself to be proficient at answering multiple-choice test questions.	Pre: ($F(1,31) = .136$, $p = .715$) No significant difference between groups Post: ($F(1,31) = .161$, $p = .691$) No significant difference between groups
Q & QX 12 In addition to content knowledge, I have strategic skills as a test-taker.	Pre: ($F(1,31) = 2.122$, $p = .155$) No significant difference between groups Post: ($F(1,31) = .471$, $p = .498$) No significant difference between groups

Null Hypothesis #2: Multiple-Choice Nursing Tests Grades

The second hypothesis of the M4Q study addressed a relationship between learning the M4Q multiple-choice test-taking strategy and scores earned on multiple-choice tests. All subjects in strata delineator 1st year (N=19) were enrolled in the same nursing courses together, were taught by the same instructors, and took the same formative and summative multiple-choice tests. All subjects in strata delineator 2nd year (N=14) were enrolled in the same nursing courses together, were taught by the same instructors, and took the same formative and summative multiple-choice tests. The groups were randomly selected by strata to maintain equality of representation in groups.

Comparisons between the groups on formative tests. Independent samples *t*-tests were calculated comparing the means of the scores on multiple-choice tests of the experimental and control groups by individual test and by types of tests. Differences between the experimental and control groups' means of scores on all formative multiple-choice test grades ($t(31) = -.821$, $p = .418$) were not statistically significant.

Using the one-way ANOVA for multiple variables to compare individual tests, no statistically significant difference was found between experimental and control group means on any of the individual formative assessments: OB1 ($F(1,31) = .509$, $p = .481$); Test 1 ($F(1,31) = .278$, $p = .602$); OB2 ($F(1,31) = .079$, $p = .781$); Test 2 ($F(1,31) = .119$, $p = .733$); OB3 ($F(1,31) = 1.069$, $p = .309$); Test 3 ($F(1,31) = 1.623$, $p = .212$); OB4 ($F(1,31) = 3.273$, $p = .080$); Test 4 ($F(1,31) = .173$, $p = .681$); Med 1 ($F(1,31) = 1.664$, $p = .252$); Med 2 ($F(1,31) = 3.246$, $p = .081$); Med 3 ($F(1,31) = 2.893$, $p = .099$); Med 4 ($F(1,31) = 1.752$, $p = .197$); Nut 1($F(1,31) = .104$, $p = .749$); Nut 2 ($F(1,31) = .339$, $p =$

.564); Demo 1 ($F(1, 31) = 1.421$, $p = .243$); Demo 2 ($F(1,31) = .012$, $p = .912$); Demo 3 ($F(1,31) = 2.005$, $p = .167$); Demo 4 ($F(1,31) = .072$, $p = .790$).

On seven of the 18 tests on which there was no statistically significant difference, the experimental group earned higher scores by an average of 2.77 percentage points. On 11 of the 18 tests, the control group earned higher scores by an average of 3.31 percentage points—a net difference between the groups of only .54 of a percentage point.

Comparisons between the groups on summative tests. Statistically significant differences were found between the two groups on their scores on (1) one of their summative tests, the HESI 2 test ($t(31) = 2.714$, $p = .011$), and (2) the means of their two summative HESI tests ($t(31) = 2.663$, $p = .012$), with the experimental group scores being statistically significantly higher than the control group scores. No significant difference was found between experimental and control group means on the instructor-prepared summative assessments entitled Final Exam ($t(31) = -.103$, $p = .919$); or on the standardized proprietary HESI 1 ($t(31) = 2.714$, $p = .287$).

Differences in HESI scores. The HESI tests are standardized, high-stakes multiple-choice tests administered with the purpose of statistically predicting the student's success on the summative program NCLEX examination for licensure, and were, therefore, the final and most important of all the tests administered in the semester. Subjects in each group took two HESI tests, the second being the final predictor test of the semester. The focus of this study was learning, practicing and applying the M4Q multiple-choice test-taking strategy. The relationships between those two variables—learning the M4Q strategy and the final HESI score, HESI 2, therefore provided a robust indication of the effectiveness of M4Q and the nursing test grades. (See Table 9.) The

difference between the experimental and control groups on the HESI 2 scores was statistically significant at the .05 criterion specified (t (31) = 2.714, p = .011), and in fact, indicated only a 1.1% chance that the difference occurred randomly.

Table 9

T-*Test for HESI 2 by Group*

Group Statistics

	Groups	N	Mean	Std. Deviation	Std. Error Mean
HESI_2	Experimental	16	88.2250	12.07358	3.01839
	Control	17	76.2765	13.14922	3.18915

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
HESI_2	Equal variances assumed	.092	.764	2.714	31	.011	11.94853	4.40277	2.96902	20.92804
	Equal variances not assumed			2.721	30.984	.011	11.94853	4.39106	2.99272	20.90434

In addition, the experimental group's scores on the HESI 2 were sufficiently higher to cause the means of the combined HESI 1 and 2 scores to also be statistically higher than the control group's scores (t (31) = 2.663, p = .012), even though the difference between the groups' earlier HESI 1 test means was not statistically significant.

CHAPTER 5: SUMMARY AND DISCUSSION

Introduction

The Summary and Discussion chapter of the dissertation briefly restates the background, problem, purpose and methodology of the research conducted for this study. The results are summarized, interpreted and discussed in the context of the literature and the theoretical framework. The study's implications are suggested, its limitations are identified, and recommendations for further studies are presented.

Statement of the Problem

Americans consider healthcare to be one of the most important social issues facing the United States. In national polls conducted between August, 2009, and March, 2010, healthcare ranked first in the category of non-economic issues identified (Saad, 2009; Bloomberg National Poll, 2010). Significant issues include the cost of care and who will provide it.

Registered nurses comprise the largest group of healthcare professionals in the United States (Buerhaus, Staiger, & Auerback, 2000). The number of nursing positions in the U.S. is expected to increase at a rate of 22% in a ten-year period from 2008 to 2018, a growth rate that is faster than the average for all occupations (Registered Nurses, 2009). There are not enough nurses available to fill the openings, however. The current shortage of professional registered nurses in the United States is unprecedented and is expected to continue (Unruh & Fottler, 2005). Addressing the need for more nurses is a multi-faceted challenge that includes factors at the national, state, institutional and individual level.

At the individual level, one aspect of the shortage is that too many individuals who want to become nurses are not successful at completing their college degrees and passing the numerous high-stakes multiple-choice tests that are required in order to achieve licensure. Each year from 2008 through the first quarter of 2010, between 10% and 14% of the examinees failed on their first attempt, a total of 33,151 potential registered nurses (NCLEX Statistics from NCSBN, 2010). An academic skill that could improve prospective nurses' ability to well-represent what they know on the requisite multiple-choice tests would offer benefit at the individual, institutional, and national level.

Numerous texts are now in print that offer advice to college students about how to maximize their performance on multiple-choice tests, but virtually all of it is based on anecdotal data and custom, rather than scientific evidence. Some of those texts are nursing-specific. For collegiate nursing instruction, formative and summative tests provide practice for students, but they are promoted and sold as statistical predictors of future success on the licensure examinations. Because the first-time pass rate on the licensure examinations is an important criterion in retaining their accreditation, college nursing programs often use multiple-choice test scores as high-stakes requirements for passing courses, progressing to the next level, and even for graduating.

The Mayfield's Four Questions (M4Q) multiple-choice test-taking strategy was developed by a college learning assistance professional over several years' practice with nursing students to help them succeed as test-takers. The literature supports the premise that one's belief in one's own ability to succeed, or perceived self-efficacy, is a key element in ultimately succeeding. Years of anecdotal evidence indicate that M4Q is

highly effective at enhancing students' ability to represent the knowledge they possess in a valid and reliable way, rather than be distracted by the distractors on the tests. This study formalized an experimental research design to test the M4Q for its impact on nursing students' perception of their own self-efficacy and their performance on instructor-prepared and standardized multiple-choice tests in nursing.

Review of the Methodology

The research design type for this experimental study was "Randomized Subjects, Pretest-Posttest Control Group Design" (Ary, Jacobs, Razavieh, & Sorensen, 2006, p. 231). Nursing students at a small Mid-western college were randomly selected by program year into an experimental group and a control group for the M4Q study. All participants signed consent forms and took a pre-intervention adapted General Self-Efficacy (GSE) (Jerusalem & Schwarzer, 1993) survey. Experimental group subjects committed to meeting with the researcher for weekly 30-minute sessions to learn and practice the M4Q strategy as the study's intervention.

Quantitative data consisted of results from pre- and post-intervention administrations of the General Self-Efficacy Scale (GSE) (Schwarzer & Jerusalem, 1993) and the participants' grades on all multiple-choice tests administered in their nursing courses during the study. Data were statistically analyzed based on an adaptation of the research design model developed by Ruban, McCoach, McGuire and Reis (2006). The statistical analyses addressed individual, class, and research group differences in subjects' responses to the 12 questions on the perceived self-efficacy survey individually and by group. It also addressed the individual and aggregate quantitative data from instructor-prepared multiple-choice test grades relative to receiving or not receiving the intervention

of learning M4Q. The data was examined for correlations between the intervention and outcomes on self-efficacy surveys and performance on multiple-choice nursing tests.

Discussion of the Conceptual Design of the Intervention

Clearly, nursing students do need test-taking skills that will prepare them to understand and correctly answer test questions written at all levels of Bloom's Taxonomy. It may not, however, be good practice—or even possible—to write multiple choice questions at all levels of Bloom's Taxonomy. While a student may have to analyze a scenario to determine appropriate action, the test-taker must only be able to recognize the correct analysis that the test writer produced. In order for the level of a question to actually be Analysis, an analysis of data, texts and/or graphics must be done by the individual taking the test. Picking the analytical answer from a list of four possibilities has a high element of chance, as well as meeting only the criterion for Comprehension or Application.

Similarly, the student's ability to synthesize cannot be assessed via a multiple choice question. By definition, Bloom's level of Synthesis requires taking discrete pieces of data or information and merging them into a new, more complete whole. The test writer might do that, and require the student test-taker to identify the correct answer, but there is no opportunity for the student to gather information into a new entity in taking a multiple choice question. Likewise, there is little or no opportunity for evaluation. Other assessment strategies are more congruent with valid assessments of the ability to analyze, synthesize and evaluate: discussion, open-ended questions, fill-ins, essays, and writing research papers. Yet, the goal persists among content-specific educators that multiple choice questions should be written at all levels.

The three lower levels of Bloom's Taxonomy—Knowledge, Comprehension and Application (Bloom, 1956), are easily and appropriately assessed with multiple-choice questions. Knowledge and Comprehension have been described as *declarative learning* (Nilson, 2003). In nursing, Knowledge includes definitions of terms, numerical equivalencies, etc. Comprehension questions demonstrate understanding of the requisite base level of Knowledge. Application questions ask the student to apply the knowledge they understand to a situation or patient. This process has been called *procedural learning* (Nilson, 2003). Nilson characterizes the three higher levels of the taxonomy—Analysis, Synthesis and Evaluation, as *reflective learning*. Nilson believes all levels can be assessed with multiple choice questions, but did not explain or demonstrate how to design questions at the higher levels.

Becoming a strategic test-taker by was supported by Bandura's Social Learning, Agency and Self-Efficacy Theories. The M4Q is a critical thinking process that was taught by one individual to another. As the learner practiced, the teacher modeled, thereby affording vicarious experiences that led to the learner's own mastery experiences. A broad range of research supports the relationships between perceived ability to succeed, or self-efficacy, and actual success. In this study, a strong sense of agency was involved, in that participation in the learning experience involved a formal commitment to participate and an ongoing commitment of time week by week. A strong statistical correlation supported the relationship between the degree of agency manifested in making and keeping the appointments with the ultimate outcomes of several of the assessments.

Summary and Discussion of Statistical Results

Conclusions Regarding Null Hypothesis 1: Self-Efficacy

The first null hypothesis of the M4Q study was, "Nursing students who were taught the Mayfield's Four Questions (M4Q) test-taking strategy will not perceive a post-intervention level of self-efficacy that is higher than the level perceived by nursing students who were not taught the strategy." The hypothesis was tested by administration of pre- and post- intervention GSE surveys. The results data were analyzed by visual analysis and statistical analysis.

Outcomes. The null hypothesis was rejected based on satisfying the 95% significance criterion based on a *t*-test for independent samples comparing the experimental and control groups' means sums of scores from pre-intervention to post-intervention GSE surveys.

Summary of survey data supporting rejection of the null hypothesis. On the post-intervention surveys, the experimental group subjects who were taught the Mayfield's Four Questions multiple-choice test-taking strategy were found to differ from the control group in the following ways:

(1) Experimental group subjects indicated higher raw scores than the control group on 11 of the 12 self-efficacy post-intervention survey statements. The group differences were statistically significant on 25% of the statements.

(2) Experimental group subjects demonstrated a statistically significant increase from the mean of pre-intervention individual subjects' sums of scores to post-intervention individual subjects' sums of scores ($t(15) = 3.94$, $p = .003$) as measured on a *t*-test for paired samples. The control group demonstrated an increase between their pre- and post-

intervention survey individual statement means from 3.44 (*sd* = .454) to 3.59 (*sd* = .662) but the increase was not statistically significant ($t(16) = -1{,}073$, $p = .299$).

Discussion. A large body of literature addressed the related concepts of self-worth, self-image, self-advocacy and, with particular application to this study, self-efficacy (Bandura, 1986; Woolfolk, Winne & Perry, 2000; Pajares, 2003; Miller & Brickman, 2004; DiRamio & Payne, 2007). All the authors found who addressed the subject agreed that there is a positive correlation between one's perception of one's own ability to succeed, and actually succeeding. For the Christian, the perceived self-efficacy emanates from a source other than one's self, and the recognition is expressed in the thought, "I can do all things through Christ, Who strengthens me." This awareness is in direct contrast to Bandura's attribution of benefits to fortuitousness.

At the most basic level, the concept of succeeding because one believes he or she can is described as self-fulfilling prophesy. In the vernacular, Henry Ford is quoted as saying, "Whether you think you can, or think you can't…you're right" (Nugent & Vitale, 2000, p. 2). Because of the correlation between self-perception and success, the results of the pre-intervention and post-intervention self-efficacy surveys in the M4Q study are important. Even if no statistical support was revealed for learning the M4Q because of improvements in grades based on this study, a statistically significant improvement in the experimental group subjects' perceived self-efficacy after learning the strategy would still have reflected a personally significant psychological gain that could manifest itself in future successes.

Null Hypothesis 2: Multiple-Choice Nursing Test Performance Results

The second null hypothesis of the M4Q study was, "Nursing students who were

taught Mayfield's Four Questions (M4Q) test-taking strategy will not earn multiple-choice test scores that are higher than the multiple-choice test scores of nursing students who were not taught the strategy." The hypothesis was tested by subjecting the percentage scores subjects earned on all the multiple-choice tests administered during the semester to statistical analysis by *t*-tests for between-group means and supported by ANOVAs.

Outcomes

The null hypothesis was rejected based on satisfying the 95% significance criterion for the differences between the experimental and control groups on

(1) *t*-tests for between-group means on the HESI 2 test, and

(2) *t*-tests for between-group means of the means of the two HESI tests.

Summary of data supporting rejection of the test scores null hypothesis. In independent samples *t*-tests, statistically significant differences were found between the experimental and control groups on their scores on the HESI 2 ($t(31) = 2.714$, $p = .011$), and the means of the two HESI tests ($t(31) = 2.663$, $p = .012$). A one-way ANOVA supported those findings, showing a significant difference between the groups on the HESI 2 ($F(1,31) = 7.365$, $p = .011$) and the mean scores ($F(1,31) = 7.094$, $p = .012$).

The supporting data is, therefore, sufficiently robust to reject both null hypotheses at the 95% criterion.

Discussion. The research strongly indicates that most formative and summative multiple-choice tests prepared and administered by nursing faculty typically are not strong assessments (Su, Osisek, Montgomery and Pellar, 2009). The questions are most frequently written at the lower levels of Bloom's Taxonomy by instructors who have had

little or no formal training in writing test questions, and contain an extraordinarily high number of test-writing errors. Yet many are used as high-stakes assessments. The test banks prepared by textbook publishers are little, if any, better (Masters, Hulsmeyer, Pike, Leichty, Miller, Verst, 2001), but most students do master the content and deal with the poorly-written questions well enough to pass. In this study, students who learned to use the M4Q were statistically more successful on some of their instructor-prepared tests than students who did not. The M4Q removes much of the risk associated with poorly-written or tricky distractors, because the distractors are not viewed as options, but are recognized for their true purpose: to distract from the correct answer. Students do not study the distractors, eliminate some, narrow them down to two, and then finally one. They decide on what the answer is based on their own usually adequate knowledge base, then skim through the options only well enough to find the correct answer, not to thoughtfully consider all the distractors.

Two large publishing houses, Elsevier and Kaplan, produce standardized practice test to prepare students for taking the NCLEX licensure examination after graduation and predict their future success based on their scores. Elsevier strongly urges that their HESI examinations only be used to assess the need for remediation, not as high-stakes tests, but HESIs are widely used as very high-stakes “gatekeeper” tests that can prevent passing course, completing a level, or even graduating. In this study, the nursing students who learned and applied the M4Q multiple-choice test-taking strategy earned statistically significant higher scores than the control group on some of their instructor-prepared tests, and perhaps more importantly, on their final HESI tests as well.

Using the strategy and doing well was not always enough motivation to continue.

Attendance patterns ranged from making appointments but virtually never keeping them to keeping virtually all appointments made. Some subjects waited for several weeks before beginning, then rarely missed. One student had been struggling academically and was at risk for not graduating because her test grades were so marginal. She learned how to use the M4Q in her first appointment. She practiced with a group on her second appointment. She practiced with only the researcher on her third appointment and did not miss any questions. She refused to believe that she could be performing so well, and announced that the researcher's practice questions must just be too easy. She chose an NCLEX review book from a shelf in the office, and asked challengingly if that could be used for practice. The researcher immediately agreed, and the student did four of the NCLEX practice questions, and got all of them right, as well. Then she never came back for another appointment.

A few weeks later, The program director announced to the faculty, in the researcher's presence, that the student described above had just passed her first HESI examination with one of the three highest scores ever earned by a student in that college. The minimum for passing was 850 points, and she had earned more than 1150. When asked how she did it, she reportedly said she had done practice test questions with a Kaplan text and the M4Q. When the students took their second HESI examination, however, the student barely passed. When asked if she had used the same strategy again, she said no, she had found a new one she decided to try. Student choices remain an uncontrolled and sometime irrational variable in experimental research.

Interpretation of the Findings

The findings of this study indicate that even rudimentary understanding and application of an evidence-based test-taking strategy can positively impact both self-efficacy perception and actual performance on multiple-choice tests. The data indicates that the M4Q strategy can positively influence both self-perception and test grades. A statistical correlation was shown between the number of M4Q practice sessions attended and test grades, indicating that continuing practice over a longer period of time may have the potential to improve test performance even more. In the anecdotal evidence from the previous nursing program, numerous students were required to learn and practice the strategy weekly for one semester as a part of a mandated remediation program, but continued to come to weekly appointments for subsequent semesters and even subsequent years, because the benefit was so apparent to them. Although this study was conducted for three full months, no subject attended more than nine sessions, yet the results indicated statistically significant gains, and supported the notion that ongoing practice for mastery leads to higher test scores.

Relationship of the Current Study to Previous Research

Three theories of Albert Bandura provided the theoretical support for this study. Bandura's Social Learning Theory defines the individual's role as an active learner. The Agency Theory aspect of Social Learning posits that the individual is an active force in his or her own learning. The Self-Efficacy Theory aspect of Agency Theory explicates the how the individual's ability to address and cope with stresses results in a belief that one can succeed, and that the belief is a factor that supports the success. Learning and applying the M4Q strategy is completely congruent with those understandings.

A significant and growing body of literature addresses the perceived self-efficacy of individuals. Emory University posted a website entitled "Student Research on Self-Efficacy: A Community of Scholars" that identified hundreds of student-designed self-efficacy studies in progress or completed by September, 2005. Although 16 studies were identified in the category *Academic Self-Efficacy*, none addressed test-taking. While 38 studies addressed *Medical Issues and Settings, Clinical, Health*, none focused on nursing students. This study made a contribution in both areas.

This study demonstrated the statistical correlations between actively learning a new strategy and the subjects' perceived self-efficacy. A significant body of literature addressed the psychological aspects of taking tests, but it almost exclusively focused on the stress and anxiety associated with test-taking (Bolger, 1990; Blankstein, Flett, & Watson, 1992; Zeidner, 2007; Thompson & Gaudreau, 2008) and the coping activities that were used or recommended to deal with stress (Bolger, 1990; Davis, DiStefano, & Schutz, 2008). A significant majority of the strategies in print that did address strategic test-taking, both in generic and nursing-specific texts, advised students to study the distractors when deciding upon a response. That was the opposite approach from the fourth step in M4Q. The published material on test-taking strategies was, with few exceptions, not supported by published research. This study made a contribution to the need for evidence-based practices in the literature.

The M4Q study made a contribution to the test-taking literature specific to nursing education in recognition of the preponderance of high-stakes multiple-choice tests required for degree completion and eventual licensure.

> A new and growing trend among nursing faculties is to set benchmarks that contain progression-to-graduation requirements based on student performance on standardized nursing exams that allow for comparison with national norms. These policies are designed to identify students who are in need of remediation prior to graduation and NCLEX candidacy so that remediation can be initiated and NCLEX failure averted. (Nibert, Young & Britt, 2003).

The most critical piece of the reporting of the HESI scores is the predictor score: the percentage of chance of passing the NCLEX on the first attempt. HESI recommends various cut-off points for progression and graduation based on these predictor scores, and a significant number of nursing programs have adopted those recommendations as policy (Nibert, Young & Britt, 2003; Spurlock, 2006). There are high stakes involved: according to a survey of programs using the HESI, the most common sanctions imposed for failure to reach the required score include inability to progress to the next semester, inability to graduate, and inability to sit for the NCLEX (Nibert, Young & Britt, 2003). Imposing even one of those consequences makes the HESI tests high-stakes. The scores on two HESI tests and the means of those scores were variables statistically analyzed in the M4Q research study. Learning the M4Q strategy was significantly correlated with the subjects' scores on their second HESI test and the means of their HESI tests, therefore contributing to the evidence-based literature related to passing the HESI examinations.

Limitations

Size

The greatest limitation of the M4Q study was the size of the sample. Based on historical patterns of persistence, previous semester enrollment data and expert opinions,

a sample of about 70 with virtually 100% participation was anticipated. Enrollment persistence was lower than expected, and agreement to participate was lower than expected. By the time consent forms were signed, only 52 participants chose to enroll in the study. Subsequent nursing department attrition, subjects' formal withdrawals and informal withdrawals due to lack of participation further reduced the samples until only 33 remained for final statistical analysis.

The small size of the nursing department meant that all 1st level students took all courses together from the same instructors, as did all 2nd level students. While that fact facilitated statistical analysis by removing the possibility of having various instructors for the same cohort be uncontrolled variables, it also gave those instructors unanticipated freedoms. Faculty have the custom of keeping classes past their scheduled time to finish teaching the content of the day, because all the students have the same next class or break between classes. When that occurred during the study, some subjects missed their meetings at the only time they could have come that week. Faculty also rearranged schedules to add study and review sessions, knowing where all the students were on any day. Students often did not come to their appointments on time or at all because of other commitments within the college that developed spontaneously. No subject came to more than 9 sessions in the three months of the study.

Another size limitation was related to number of intervention personnel. The researcher was the only teacher of the intervention strategy. While it facilitated scheduling, having only one instructor could have brought in intangible variables such as personality fit that impacted the degree of voluntary participation.

Institutional Climate

Another limitation to the M4Q study was the institutional climate. The foundational value of nursing practice is caring, and the faculty, staff and students manifested that value in numerous observable ways. The entire nursing department is contained in three large classrooms, one of which also serves as a skills lab and a faculty lounge; a department office; four individual faculty offices; and one three-desk adjunct faculty office. These spaces are all adjacent to one another in the basement of an academic building. Most of the students, faculty and staff know each other by name and sight and see each other many times each week. That level of familiarity made maintaining anonymity regarding who was in the experimental group and who was in the control group impossible. Anyone who met with the researcher was observed doing so by a number of people, and anyone using the M4Q card to cover distractors could be seen by many.

Recommendations for Further Research

This M4Q experimental study was one of very few that addressed specific learning strategies, and even fewer that addressed test-taking strategies. In light of the ubiquitousness of high-stakes testing at every level of American education, formal experimental testing of new strategies that anecdotally appear to work should become the norm, rather than the exception. The NCLEX failure rate for nursing licensure is a concrete demonstration of the need to find new test-taking strategies for undergraduate collegiate nursing education. Every candidate for licensure has already mastered the nursing curriculum content at a level sufficient to complete a degree program and

graduate from an accredited college, so a deficiency great enough to result in failure might reasonably be assumed to be related to high-stake test-taking skills.

The M4Q multiple-choice test-taking strategy was successfully used in one setting with baccalaureate nursing students at all four levels of the program, but only anecdotal evidence was collected. It has now been experimentally tested in another setting in which the subjects were again nursing students, but who were first- and second- year students in an associate degree program. The samples were large enough to observe some statistical significance, but were small. The M4Q should be tested on larger populations of nursing students and populations of students who are not nursing majors to see if statistically significant benefits result for them, as well.

References

Alfaro-LeFevre, R. (1995). *Critical thinking in nursing: A practical approach*. Philadelphia: W. B. Saunders.

Alfaro-LeFevre, R. (2009). *Critical thinking and clinical judgment: A practical approach to outcome-focused thinking.* St. Louis: Saunders Elsevier.

American Educational Research Association (2000). AERA Position Statement on High-Stakes Testing in Pre-K-12 Education (2000). Retrieved from http://www.aera.net/Printable.aspx?id=378

American Psychological Association (2007). APA Online public affairs. Appropriate Use of High-Stakes Testing in Our Nation's Schools. Retrieved from http://www.apa.org/pubsinfo/brochures/testing.aspx

Ary, D., Jacobs, L. C., Razavieh, A., Sorensen, C. (2006). Introduction to research in education. Belmont, CA: Thomson Higher Education.

Bandura, A. (1967). Behavioral modifications through modeling procedures. In L. Krasner and L. Ullmann (Eds.), *Research in behavior modification: New developments and implications* (pp. 310-340). New York: Holt, Rinehart and Winston.

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice Hall. In F. Pajares (Ed.), Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly, 19*, 139-158. doi: 10.1080/10573560390143085

Bandura, A. (2002). Social cognitive theory in cultural context. *Applied Psychology, 31*(2), 269-290. Retrieved from

http://web.ebscohost.com.ezproxy.liberty.edu:2048/ehost/pdfviewer/pdfviewer?vid=5&hid=9&sid=aec63ac-5f16-4722-8973-c2a83e377f8a%40sessionmgr12

Bandura, A. (2006a). Guide for constructing self-efficacy scales. In F. Pajares & Urdan, T. (Eds.) *Adolescence and Education Series: Vol. 5. Self-Efficacy Beliefs of Adolescents*. (pp. 307-337). Retrieved from http://des.emory.edu/mfp/014-BanduraGuide2006.pdf and http://www.des.emory.edu/mfp/AdoEd5html

Bandura, A. (2006b). Toward a psychology of human agency. *Educational Psychologist, 1*(2), 164-180. Retrieved from http://des.emory.edu/mfp/Bandura2006PPS.pdf

Bandura, A. (2008). Reconstrual of "free will" from the agentic perspective of social cognitive theory. In J. Baer, J. C. Kaufman & R. F. Baumeister (Eds.), *Are we free? Psychology and free will* (pp. 86-127). Oxford: Oxford University Press. Retrieved from http://des.emory.edu.mfp/Bandura2008AgencyFreeWill.pdf and http://www.oup.com/us/

Beatrice, J. A. (1995). *Learning to study through critical thinking*. Chicago: Richard D. Irwin.

Blankstein, K. R., Flett, G. L., & Watson, M. S. (1992). Coping and academic problem-solving ability in test anxiety. *Journal of Clinical Psychology, 48*(1), 37-47. Retrieved from http://web.ebscohost.com.ezproxy.liberty.edu:2048/ehost/pdfviewer?vid=4&hid9&sid=1568f125-c134-4d56-b0f1-cbc9e58b6b1e%40sessionmgr14

Bloom, B. S. (Ed.) (1956). Taxonomy of educational objectives: The classification of educational goals. New York: McKay.

Bloomberg Poll (2010, March 19-22). Problems and priorities. PollingReport.com. Retrieved from http://pollingreport.com/prioriti.htm

Boekaerts, M. (1993). Being concerned with well-being and with learning. *Educational Psychologist 28*(2), 149-167. Retrieved from http://www.questia.com/library/jp-educational-psychologist.jsp?CRID=jp_educational_psychologist&OFFID =se12&KEY=educational psychologist

Bolger, N. (1990). Coping as a personality process: A prospective study. *Journal of Personality and Social Psychology, 59*(3), 525-537. Retrieved from http://web.ebscohost.com.ezproxy.liberty.edu:2048/ehost/pdfviewer?vid=11&hid=9&sid=1568f125-c134-4d56-b0f1-cbc9e58b6b1e%40sessionmgr14 and www.apa.org

Buerhaus, P. I., Staiger, D. O., & Auerback, D. I. (2000). Implications of an aging registered nurse workforce. *Journal of the American Medical Association*, 283, 2948-2954. Retrieved from http://jama.ama-assn.org

Carpenter, S. (May, 2001). The high stakes of educational testing. Monitor on Psychology. Retrieved from http://www.apa.org/monitor/may01/edtesting.aspx

Chenovert, M. (2006*). Mosby's tour guide to nursing school: A student's road survival kit.* Saint Louis: Mosby Elsevier.

Cohen, A. D. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly, 3*(4), 307-331. Retrieved from http://www.routledge.com

Cohen, M., Ben-Zur, H., & Rosenfeld, M. J. (2008). Sense of coherence, coping strategies, and test anxiety as predictors of test performance among college

students. *International Journal of Stress Management, 15*(3), 289-303. Retrieved from http://www.stress-management-nl.org/journal.html

Davis, H. A., DiStefano, C., Schutz, P. A. (2008). Identifying patterns of appraising tests in first-year college students: Implications for anxiety and emotion regulation during test taking. *Journal of Educational Psychology, 100*(4), 942-960. Retrieved from apa.org

Dembo, M. H. (2004). *Motivation and learning strategies for college success: A self-management approach.* Mahwah, N. J.: Lawrence Erlbaum Associates.

DiRamio, D., & Payne, R. (2007). Assessing the relationship between campus programs, student self-efficacy, stress, and substance abuse. *College Student Journal, 41*, 676-693. Retrieved from http://www.projectinnovation.biz/csj_2006.html

DiscoverNursing.com (2010). Johnson & Johnson. Retrieved from http:/www.discovernursing.com/jandj.aspx

Dodeen, H. (2008). Assessing test-taking strategies of university students: Developing a scale and estimating its psychometric indices. *Assessment & Evaluation in Higher Education, 33*(4), 409-419. Retrieved from http://ejournals.ebsco.com/Issue.asp?IssueID=813557

Downey, T. A. (2006). Indicia: Letters to the Editor. *Computers, informatics, nursing & nurse educator.* Retrieved from http://www.cinjournal.com

Downing, S. (2005). *On course: Strategies for creating success in college and in life*. Boston: Houghton Mifflin.

Elsevier (2007). Heritage. Retrieved from http://www.elsevier.com/wps/find/intro.cws_home/heritage

Elsevier: Mosby's AssessTest (2008). Frequently asked questions. Retrieved from http://www1.us.elsevierhealth.com/OnlineAssessTest/faqs.html

Evolve: Instructor site: Evolve Reach (2008). Remediation. Retrieved from http://evolve.elsevier.com/staticPages/I_reach_exams_summary2.html

Evolve: Student site: Evolve Reach (2008). The HESI experts. Retrieved March 3, 2008, from http://evolve.elsevier.com/staticPages/s_reach_history_experts.html

Ferrett, S. K. (1996). *Connections: Study skills for college and career success*. Chicago: Irwin Mirror.

Ferrett, S. K. (2010). *Peak performance: Success in college and beyond*. Boston: McGraw Hill.

Field, A. (2005). *Discovering statistics using SPSS: (and sex, drugs, and rock 'n roll).* London: Sage Publications.

Fox News Poll/Opinion Dynamics (April 8, 2010). Retrieved from http://www.foxnews.com/projects/pdf/040810_Obama_HC_2010_web.pdf

Fraenkel, J. R. & Wallen, N. E. (2003). *How to design and evaluate research in education*. Boston: McGraw Hill.

Frith, K. H., Sewell, J. P., & Clark, D. J. (2005). Best practices in NCLEX-RN readiness preparation for baccalaureate student success. *CIN: Computers, informatics, nursing*, *23*,6, 322-329, reprinted in *CIN: Computers, informatics, nursing & nurse educator*, May/June, 2006 [Suppl.]. Retrieved from http://www.cinjournal.com

Gardyasz, J. (2010, March 6). Coalition sees to address Iowa's nurse shortage. *Business Record.* Retrieved from http://www.businessrecord.com/main.asp?SectionID=5&SubSectionID=10&ArticleID=9679

Goodin, H. J. (2003). The nursing shortage in the United States of America: An integrative review of the literature. *Journal of Advanced Nursing, 43*(4), 335-350. Retrieved from http://www.journalofadvancednursing.com/docs/1365-2648.2003.02722-1.x.pdf

Grusec, J. E. (1999). Social learning theory and developmental psychology: The legacies of Robert Sears and Albert Bandura. In R. Parke, P. Ornstein, P. Rieser, J. John, & C. Zahn-Waxler (Eds.), *A century of developmental psychology* (pp. 473-497). Washington: American Psychoogical Association.

Hanks, C. (2006). Indicia: Letters to the Editor. *Computers, informatics, nursing & nurse educator.* May/June, 2006[Suppl.] Retrieved from http://www.cinjournal.com

Hoefler, P. A. (1995). *Successful problem-solving & test-taking for the beginning nursing student.* Silver Spring, MD: Medical Education Services.

Illinois State Board of Education (n. d.). No child left behind (NCLB). Retrieved from http://www.isbe.state.il.us/nclb/default.htm

Iowa needs nurses now. Retrieved from http://www.ianeedsnursesnow.org

Jerusalem, M., & Schwarzer, R. (1979). The General Self-Efficacy Scale (GSE) [Assessment instrument]. Retrieved from http://userpage.fu-berlin.de/~health/selfscal.htm

Judson, E. (2007). Retaking a high stakes mathematics test: Examination of school interventions and environments. *American Secondary Education 36*(1), 15-30. Retrieved from http://www.ashland.edu/ase

Kanar, C. C. (2004). *The confident student*. Boston: Houghton Mifflin.

Kaplan Nursing (2008). Integrated Testing for Nursing Schools. [Web page]. Retrieved from http://www.kaplannursing.com

Katz, J. R. (2004*). Keys to nursing success*. Upper Saddle River, NJ: Pearson Prentice Hall.

Langan, J. (2001). *Ten skills you really need to succeed in school*. West Berlin, NJ: Townsend.

Lazarus, R. S., & Folkman, S. (1984). Stress, appraisal, and coping. New York: Springer.

Lippke, S., Wiedemann, A. U., Ziegelmann, J. P., & Schwarzer, R. (2009). Self-efficacy moderates the mediation of intentions into behavior via plans. *American Journal of Health Behavior, 33*(5), 521-529. Retrieved from http://www.ajhb.org/

Longo, D., & Lent, R. (1992). Social cognitive variables in the prediction of client motivation and attrition. *Journal of Counseling Psychology* [Serial online], *39*, 447. Retrieved from www.apa.org.

Luskin, C. V. (2000). *A magical place*. Stanford University Bing Nursery School [web site]. Retrieved from http://www.stanford.edu/dept/bingschool/aboutbing_gentry.html

Martin, J. (2004). Self-regulated learning, social cognitive theory, and agency. *Educational Psychologist, 39*, 133-145. Retrieved from www.apa.org

Masters, J. C., Hulsmeyer, B. S., Pike, M. E., Leichty, K., Miller, M. T., Verst, A. L. (2001). Assessment of multiple-choice questions in selected text banks accompanying text books used in nursing education. *Journal of Nursing Education, 40*(1), 25-32. Retrieved from http://www.journalofnursingeducation.com

Massie, J. L. (1971). Essentials of management. Englewood Cliffs, NJ: Prentice-Hall.

Miller, R. B., & Brickman, S. J. (2004). A model of future-oriented motivation and self-regulation. *Educational Psychology Review, 16*(1), 9-33. Retrieved from http://www.springer.com/psychology/journal

Morrison, S. M., Adamson, C., Nibert, A., & Hsia, S. (2004). HESI exams: An overview of reliability and validity. *CIN: Computers, Informatics, Nursing, 22*, 4, 220-226. Retrieved from http://journals.lww.com/cinjournal/pages/default.aspx

NCLEX statistics from NCSBN (2010). Number of candidates taking NCLEX examination and percent passing, by type of candidate. Retrieved from https://www.ncsbn.org/Table_of_Pass_Rates_2010.pdf

NCLEX statistics from NCSBN (2009). Number of candidates taking NCLEX examination and percent passing, by type of candidate. Retrieved from https://www.ncsbn.org/Table_of_Pass_Rates_2009.pdf

NCLEX statistics from NCSBN (2008). Number of candidates taking NCLEX examination and percent passing, by type of candidate. Retrieved from https://www.ncsbn.org/Table_of_Pass_Rates_2008.pdf

Nibert, A. T., Young, A. (2001). A third study on predicting NCLEX success with the HESI exit exam. *Computers in Nursing, 19*(4), 173-178, reprinted in *CIN:*

Computers, informatics, nursing & nurse educator, May/June, 2006[Suppl. 3], 21S-27S. Retrieved from http://www.cinjournal.com to http:/www.nursingcenter.com/library/JournalArticle.asp?Article_ID_648187 and http://journals.lww.com/conjournal/pages/default.aspx

Nibert, A. T., Young, A., & Adamson, C. (2002). Predicting NCLEX success with the HESI Exit Exam: Fourth annual validity study. *Nurse Educator*, *28*(3), reprinted in *CIN: Computers, informatics, nursing & nurse educator*, May/June, 2006[Suppl. 3], pp. 28S-34S. Retrieved March 7, 2008, from http://www.cinjournal.com to http://www.nursingcenter.com/library/JournalArticle.asp?Article_ID_648190 and http://journals.lww.com/cinjournal/pages/default.aspx

Nibert, A. T., Young, A., Britt, R. (2003) The HESI exit exam: Progression benchmark and remediation guide. *Nurse Educator*, *28*(3), pp. 141-145, reprinted in *CIN: Computers, informatics, nursing & nurse educator*, May/June, 2006[Suppl.], 57S-61S. Retrieved from http://www.cinjournal.com

Nilson, L. (2003). *Teaching at its best: A research-based resource for college instructors*. Bolton, MA: Anker.

Nursing Shortage Fact Sheet (2009). American Association of Colleges of Nursing. Retrieved September 21, 2009, from www.aacn.nche.edu/Media/shortageresource.htm

Nursing Shortage Fact Sheet (2010). American Association of Colleges of Nursing. Retrieved from http://www.aacn.nche.edu/Media/pdf/NrsgShortageFS.pdf

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66,* 543-578. Retrieved from and http://rer.sagepub.com/

Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & writing quarterly, 19*(1), 139-158. Retrieved from http://www.tandf.co.uk.ezproxy.liberty.edu:2048/journals/tf/10573569.html

Pajares, F. (2004). *Albert Bandura: Biographical sketch*. Retrieved from http://des.emory.edu/mfp/bandurabio.html

Powers, W. T. (1991). Commentary of Bandura's 'human agency.' *American Psychologist, 46*(2), 151-153. Retrieved from www.apa.org

RCMAR Measurement Tools (2006). Generalized Self-Efficacy Scale (GSE). SC Cooperative for Healthy Aging in Minority Populations: Resource Centers for Minority Aging Research. Retrieved from http://www.musc.edu/dfm/RCMAR/GenSEscale.html

Reed Elsevier (2008). Press releases: Science & medical. Health Education Systems Incorporated (HESI) joins Elsevier. Retrieved March 3, 2008, from http://www.reed-elsevier.com/index.cfm?Articleid=1616

Registered Nurses (2009). *Occupational Outlook Handbook, 2010-11 Edition*. Bureau of Labor Statistics. Retrieved from http://www.bls.gov/oco/ocos083.htm.

Renner, B., Spivak, Y, Kwon, S., & Schwarzer, R. (2007). Does age make a difference? Predicting physical activity of South Koreans. *Psychology and Aging, 22*(3), 482-493. Retrieved from www.apa.org

Reynolds, J. A. (1996*). Succeeding in college: Study skills and strategies*. Boston: Allyn & Bacon.

Risser, J., Jacobson, T. A., & Kripalani, S. (2007). Development and psychometric evaluation of the Self-Efficacy for Appropriate Medication Use Scale (SEAMS) in low-literacy patients with chronic disease. *Journal of Nursing Measurement, 13*, 203-219. Retrieved from http://www.springerpub.com/product/10613749

Rottschaefer, W. A. (1991). Some philosophical implications of Bandura's social cognitive theory of human agency. *American Psychologist, 46*(2), 153. Retrieved from www.apa.org

Rachal, K. C., Daigle, S., & Rachal, W. S. (2007). Learning problems reported by college students: Are they using learning strategies? *Journal of Instructional Psychology, 34*(4), 191-199. Retrieved from http://www.projectinnovation.biz/index.html

Rubin, L. M., McCoach, D. B., McGuire, J. M., & Reis, S. M. (2003). The differential impact of academic self-regulatory methods on academic achievement among university students with and without learning disabilities. *Journal of Learning Disabilities, 36*(3), 270-286. Retrieved from http://ldx.sagepub.com

Saad, L. (2009, October 14). U.S. satisfaction sinks to six-month low: Economy remains nation's most important problem, but mentions of it are down. Gallup. Retrieved from http://www.gallup.com/poll/123662/U.S.-Satisfaction-Sinks-Six-Month-Low.aspx?version=print

Scherbaum, C. A., Cohen-Charash, Y., & Kern, M. J. (2006). Measuring general self efficacy: A comparison of three measures using item response theory. *Educational and Psychological Measurement, 66*(6). Retrieved from http://epm.sagepub.com

Schwarzer, R. (2009). General Self-Efficacy Scale. Retrieved from http://userpage.fu-berlin.de/~health/selfscal.htm

Schwarzer, R, & Jerusalem, M. (1993). General Self-Efficacy Scale. Retrieved from http://userpage.fu-berlin.de/~health/engscal.htm

Sides, M. B. & Korchek, N. (1998*). Successful test-taking: Learning strategies for nurses*. Philadelphia: Lippincott.

Spurlock, D. (2006). Do no harm: Progression policies and high-stakes testing in nursing education. *Journal of nursing education*, *45*(8), 297-302. Retrieved from http://www.journalofnursingeducation.com/

Swartz, S. M. (2006). Acceptance and accuracy of multiple choice, confidence-level, and essay question formats for graduate students. *Journal of Education for Business, 81*, 215-220. Retrieved from http://www.heldref.org/pubs/jeb/about.html

The 2007 State of America's Hospitals—Taking the Pulse (2007). American Hospital Association. Retrieved from www.ahapolicyforum.org/

U.S. Department of Labor, Bureau of Labor Statistics (2009). *Occupational outlook handbook, 2010-2011 Edition*, Registered Nurses. Retreived from http://www.bls.gov/oco/ocos083.htm.

Welkowitz, J., Ewen, R. B., & Cohen, J. (2000). *Introductory statistics for the behavioral sciences.* Fort Worth: Harcourt Brace College.

Wong, L. (2006). *Essential study skills*. Boston: Houghton Mifflin.

Wood, D. (2010). Why have nurses left the profession? Nursezone.com [website]. Retrieved from http://www.nursezone.com/nursing-news-events/more-news/Why-Have-Nurses-Left-the-Profession_29118.aspx

Woolfolk, A. F., Winne, P. H., & Perry, N. F. (2000). *Educational psychology* (Canadian ed.) Scarborough, Ontario, Canada: Allyn and Bacon, Canada. In J. Martin (2004). Self-regulated learning, social cognitive theory, and agency. *Educational Psychologist, 39*(2), 133-145. Retrieved from www.apa.org

Zeidner, M. (1995). Coping with examination stress: Resources, strategies, outcomes. *Anxiety, Stress, and Coping, 8*, 279-298. Retrieved from http://www.routledge.com

Zeidner, M. (2007). Test anxiety in education contexts: Concepts, findings, and future directions. In P. A. Schutz, R. Pekrun, & G. D. Reinhard (Eds.), *Emotion in Education*, 165-184. Boston: Elsevier Academic. Retrieved from http://www.elsevier.com/wps/find/homepage.cws_home

Appendix A

IRB Approval 786.010810: Effect of the Mayfield¹s Four Questions (M4Q) Test-taking Strategy on Nursing Students¹ Perceived Self-efficacy and Multiple Choice Test Scores

Milacci, Ellen Elizabeth

Sent: Friday, January 29, 2010 11:47 PM

To: Mayfield, Linda Riggs; lmayfield@brcn.edu

Cc: Parker, Karen (School of Education); Garzon, Fernando L.; Institution Review Board

Attachments: _____)[Open as Web Page]; _____)[Open as Web Page]

Dear Linda,

We are pleased to inform you that your above study has been approved by the Liberty IRB. This approval is extended to you for one year. If data collection proceeds past one year, or if you make changes in the methodology as it pertains to human subjects, you must submit an appropriate update form to the IRB. Attached you'll find the forms for those cases.

Thank you for your cooperation with the IRB and we wish you well with your research project. We will be glad to send you a written memo from the Liberty IRB, as needed, upon request.

Sincerely,

Fernando Garzon, Psy.D.
IRB Chair, Liberty University
Center for Counseling and Family Studies Liberty University
1971 University Boulevard
Lynchburg, VA 24502-2269
(434) 592-4054
Fax: (434) 522-0477

Liberty University IRB Approval

Appendix B

Blessing-Rieman College of Nursing IRB Approval

Blessing-Rieman College of Nursing
Institutional Review Board (IRB)

January 25, 2010

Linda Riggs Mayfield MA
Blessing-Rieman College of Nursing
Broadway @ 11th
Quincy, IL 62305-7005

RE: Effect of the Mayfield's Four Questions (M4Q) Test-taking Strategy on the Perceived Self-efficacy and Multiple Choice Test Grades of Nursing Students.

Dear Ms. Mayfield:

Your study: Effect of the Mayfield's Four Questions (M4Q) Test-taking Strategy on the Perceived Self-efficacy and Multiple Choice Test Grades of Nursing Students has been granted IRB approval. Your proposal qualified for an expedited review. The research proposal submitted meets the criteria and requirement of minimal risk to participants as outlined in IRB policy and procedures. Your study has been assigned the following IRB number IRB # 2010-2.

The committee finds the protection of human subjects is adequately outlined and accounted for in the description of the research. A change in research protocol should be obtained from the IRB website if you are considering changing the submitted study design. You should file an adverse event form immediately with the IRB office if such an event occurs during your study.

There is no need for interim reports during the study if it is completed by the projected end date. If the study is not completed on time, an Extension Form should be filed with the IRB one month prior to August 31, 2010. Please submit a final report to the IRB Committee after the study is completed and the analysis of data has been done.

Sincerely,

Carol Ann Moseley

Carol Ann Moseley, RN, PhD
Chair, Institutional Review Board

Approved 10/16/09

Appendix C

Informed Consent Form

CONSENT FORM

Effect of the Mayfield's Four Questions (M4Q) Test-taking Strategy

on Nursing Students' Perceived Self-efficacy

and Multiple Choice Test Scores

Linda Riggs Mayfield, MA

Your Invitation

You are invited to be in a research study of a test-taking strategy for multiple-choice questions. You were selected as a possible participant because you are a nursing student at Hannibal-LaGrange College (H-LG). All the basic track first and second year nursing students in the associate degree plan are invited to be participants. I ask that you read this form and ask any questions you may have before signing your name to agree to be in the study.

This study is being conducted by: Linda Riggs Mayfield, MA, doctoral student, Liberty University, Lynchburg, VA; and associate faculty for academic enhancement at Blessing-Rieman College of Nursing (B-RCN), Quincy, IL.

Background Information

The purpose of this study is to determine if nursing students who are taught the Mayfield's Four Questions (M4Q) test-taking strategy (1) think of themselves as better test-takers, and/or (2) earn higher scores on multiple-choice tests than nursing students who are not taught M4Q.

Procedures:

If you agree to be in this study, you commit to do the following things:

(1) Become a member of either the control group or the experimental group. There will be a random drawing from all students who sign the Consent Form, to determine whether each one will be in the experimental group or the control group. The drawing will place freshmen and sophomores in the groups as equally as possible. Every participant has an equal chance of being chosen for the experimental group or the control group.
(2) Take a short survey--12 questions on a Likert scale, at the beginning of the study and end of the study. The study will be conducted during the spring semester of the 2009-

2010 academic year. Your name will be on the survey so your answers can be recorded for the correct group, but your answers will be coded and recorded for statistical analysis according to the group you are in, not as an individual, so your confidentiality is protected.

(3) Agree to allow your grades on multiple choice questions in tests in your nursing courses for the spring semester to be collected and statistically analyzed as part of your group. Your nursing faculty will send your test grades to the H-LG Nursing Office secretary, and she will record them according to your group, and send the scores to me for statistical analysis at the end of the semester. The test grades will be identified by a coded number the secretary assigns, not your name. I will not know which grades were earned by which individual students, so your confidentiality is protected. A separate list of the names and their confidential coded identity will be created by the H-LG Nursing Office secretary, to be kept with the data, in case the participants ever need to be identified. The survey results and test grades will be archived separately from the document with the participants' names.

Research rules require that the data from the study is kept for at least seven years before it is destroyed. There is no plan for ever destroying the data. It will be stored electronically and in hard copy and protected in exactly the same way that other personally-identifiable data is kept at H-LG and B-RCN. Your privacy rights are protected by a federal law called FERPA.

(4) Agree to abide by the research design for your group, the experimental group or the control group. Mixing the experiences of the two groups compromises the results of the research.

Students who are randomly chosen to participate in the control group will have their regular educational experience this semester. If you are in the control group, you will agree not to talk to faculty or students in the experimental group about the intervention.

Students randomly chosen to participate in the experimental group will receive the "intervention" of also being taught something new. They will meet with me, the researcher, at least once with the whole experimental group, to learn the details about the project, then each individual will meet with me weekly, as part of a small group, for a teaching/learning session, at a place and time they help choose. If you are in the experimental group, you will agree not to talk about the intervention with your faculty or with students in the control group.

Risks and Benefits of Being in the Study

The study has some risks: (1) There is a risk that you will be chosen for the experimental group and invest your time in the study, but experience no improvement as a test-taker. (1) There is the risk that students in the control group and experimental group will have a significant difference in their test scores and grades; but there is no way to know which group would be higher. (2) There is the risk that you are in the experimental group, and learn the strategy, but using it makes you do worse on your tests. This risk is highly unlikely, because the

strategy has already been used for many years in another college and has been taught there in several different ways. The intervention will only be conducted in a manner that was successful in the past. (3) There is a risk that your perceived self-efficacy and test scores not remain confidential. This risk is also very minimal. All those kinds of data already exist in your college, and in the college in which the researcher teaches, and your right to confidentiality is protected by law. Your college's Nursing Office staff, and the researcher and the clerical worker(s) in her office are used to maintaining complete confidentiality, and are aware that there can be severe penalties if confidentiality is not kept.

The benefits to participation are: (1) If you are randomly chosen for the experimental group, and the intervention helps you, you might find that you see yourself as a better test-taker, and/or you might do better on multiple choice tests. (2) If you are chosen for the control group, you will have all the benefits you normally would at H-LG, and if the M4Q does benefit the experimental group, you will be taught it after the research data is collected. If it does not help, you will not have lost anything. (3) You will have the educational experience of participating in a formal, Institutional Review Board-approved research study. You may cite that in a resume or job application.

Compensation:

You will receive no payment for participating.

Confidentiality:

The records of this study will be kept private. In any kind of report or article which might be published, no personally-identifiable information will be included. Research records will be stored securely, both electronically and in hard copy, only researchers and their clerical worker(s) will have access to the records, and all your confidentiality rights under FERPA will be respected. There is no plan to destroy the data when this study is complete, and they might be used for additional research studies in the future. If so, all confidentiality rights continue.

Voluntary Nature of the Study:

Participation in this study is completely voluntary. Your decision whether or not to participate will not affect your current or future relations with the Liberty University, Blessing-Rieman College of Nursing, Hannibal-LaGrange College, or your course instructors. If you decide to participate, you are free to not answer any question, not participate in any session, and withdraw at any time without affecting those relationships or suffering any consequences.

Contacts and Questions:

The researcher conducting this study is: Linda Riggs Mayfield, MA. You may ask any questions you have now. If you have questions later, you are encouraged to contact me at Blessing-Rieman College of Nursing, 217-228-5520 x 6997, or lmayfield@brcn.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher, you are encouraged to contact the chairperson of the Liberty University Institutional Review Board, Dr. Fernando Garzon, Chair, 1971 University Blvd, Suite 2400, Lynchburg, VA 24502, fgarzon@liberty.edu; or the chairperson of the Blessing-Rieman College of Nursing Institutional Review Board, Dr. Carol A. Moseley, Broadway @ 11th Street, Quincy, IL, 62305-7005, 217-228-5520, cmoseley@brcn.edu.

You will be given a copy of your signed Consent Form to keep for your records. A copy will be kept with the research data. The director of your nursing program will also keep a copy on file, and your nursing faculty members will be notified that you have given permission for your test grades to be disclosed to the researcher as part of your group's grades.

Statement of Consent:

I have read the above information. The research study has been explained and I have been given the opportunity to ask questions. I consent to participate in the study.

PRINTED NAME:_____

Signature:_____Date: _____

Signature of Researcher:_____ Date: _____

Appendix D

Pre-and Post-Intervention General Self-Efficacy (GSE) Survey

Effect of the Mayfield's Four Questions (M4Q) Test-taking Strategy

on Nursing Students' Perceived Self-efficacy

and Multiple Choice Test Scores

Pre- and Post-Intervention General Self-Efficacy (GSE) Survey

Research Study: Effect of the Mayfield's Four Questions (M4Q) Test-taking Strategy on Nursing Students' Perceived Self-Efficacy and Multiple Choice Test Scores

Researcher: Linda Riggs Mayfield, MA
Associate Faculty, Blessing-Rieman College of Nursing, Quincy, IL 62305
Doctoral Student, Liberty University, Lynchburg, VA 24502

Self-Efficacy Pre-Intervention Survey

Participant's Name:_____Date: _____

Your name will not be associated with the data from your survey in the statistical analysis or any reporting of the results. Your data will become part of your group's data. Circle the number on the Likert Scale that you think best represents your level of agreement with each statement. These questions are only descriptive; there are no right or wrong answers.

1. I can always manage to solve difficult problems if I try hard enough.

1	2	3	4	5
Not at all true				Completely true

2. If someone opposes me, I can find the means and ways to get what I want.

1	2	3	4	5
Not at all true				Completely true

3. It is easy for me to stick to my aims and accomplish my goals.

1	2	3	4	5
Not at all true				Completely true

4. I am confident that I could deal efficiently with unexpected events.

1 2 3 4 5
Not at all true Completely true

5. Thanks to my resourcefulness, I know how to handle unforeseen situations.
1 2 3 4 5
Not at all true Completely true

6. I can solve most problems if I invest the necessary effort.
1 2 3 4 5
Not at all true Completely true

7. I can remain calm when facing difficulties because I can rely on my coping abilities.
1 2 3 4 5
Not at all true Completely true

8. When I am confronted with a problem, I can usually find several solutions.
1 2 3 4 5
Not at all true Completely true

9. If I am in trouble, I can usually think of a solution.
1 2 3 4 5
Not at all true Completely true

10. I can usually handle whatever comes my way.
1 2 3 4 5
Not at all true Completely true

11. I consider myself to be proficient at answering multiple-choice test questions.
1 2 3 4 5
Not at all true Completely true

12. In addition to content knowledge, I have strategic skills as a test-taker.
1 2 3 4 5
Not at all true Completely true

1-10. Ralf Schwarzer & Matthias Jerusalem, 1993. Used by permission.

Appendix E

Handout for Mayfield's Four Questions (M4Q)

Mayfield's Four Questions© (M4Q)

Linda Riggs Mayfield, MA ©2007

This research-based multiple-choice test-taking strategy decreases the power of a distracter to distract the person who knows the correct answer. To maximize this strategy, cover up all the possible answers with a card or paper containing the Mayfield's Four Questions (M4Q), then work through the M4Q questions to find the answer to the test question. Do not look at the options until you have gone through all four M4Q questions about that test question.

1. What is this question really asking me?
   Identify the important words. If the item is not set up as a question, use all the important words to rephrase it as a question.

2. What is it trying to determine that I know?
   This is not the same as 1, above. The question might be trying to determine if you know priorities, the meanings of certain terms, universal precautions, therapeutic communication, or basic anatomy, for example.

3. What level of learning is needed to answer correctly?
   See Bloom's Taxonomy, below. Research conducted on published test banks of nursing questions indicates the vast majority of multiple choice questions are application level or below. Knowing this gives the test-taker confidence, and also helps you not read into the question.

4. What do I think is the correct answer?
   If you determine what you think is the correct answer before you look at the options, you do not need to carefully read and study all those distractors—careful reading and study increases the opportunity for them to appear to be correct. Just skim through them, looking for what you decided was the correct answer. Unless there is an umbrella statement*, go with it.

The Levels of Learning, Based on Bloom's Taxonomy:

Knowledge is the lowest, most basic, concrete level of learning, followed by comprehension of the knowledge and the ability to apply the knowledge. Multiple choice test question format is good for determining a student's knowledge, comprehension and ability to apply the knowledge they comprehend. The higher levels are usually tested by a different format—discussion, writing, research.

Evaluation

Synthesis

Analysis

Application

Comprehension

Knowledge (Information)

Umbrella Statements

Umbrella statements are true statements that are broader, or more inclusive, than the distracters. They cover more situations. If you find that more than one of the options is true, look for the one that is broader, or covers more possible situations—an Umbrella Statement.