

To the Graduate Council:

I am submitting herewith a dissertation written by Barbara R. Norwood entitled "Critical Thinking in Nursing Students: A Comparison of Designs." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Education, with a major in Leadership Studies in Education.

Norma T. Mertz

Norma T. Mertz, Ed.D.
Major Professor

We have read this dissertation
and recommend its acceptance:

Jeffery P. Aper
Jeffery P. Aper, Ph.D.

Mary Jane Connelly
Mary Jane Connelly, Ed.D.

Dr. Linda Pickthorne Fletcher
Dr. Linda Pickthorne Fletcher, Ph.D.

Accepted for the Council:

C. W. Minkel

C. W. Minkel, Ph.D.
Associate Vice Chancellor and
Dean of the Graduate School

CRITICAL THINKING IN NURSING STUDENTS:
A COMPARISON OF DESIGNS

A Dissertation
Presented for the
Doctor of Education
Degree
The University of Tennessee, Knoxville

Barbara R. Norwood
May 1998

Copyright © Barbara R. Norwood, 1998

All rights reserved

DEDICATION

To the memory of my mother, Margaret Seaton Rohrer, who instilled in me the belief
that I could do anything I set my mind to do

ACKNOWLEDGMENTS

There have been many people who have provided support and guidance throughout this endeavor. First and foremost, I owe a great debt of gratitude to my committee members. I would like to thank Dr. Linda Fletcher, whose cheerfulness and willingness to travel for committee meetings will always be appreciated; Dr. Mary Jane Connelly, who was supportive and knew when a little humor would help. It was a privilege to work with Dr. Jeffery Aper who constantly encouraged me to stretch and look beyond the obvious and without whose encouragement and help this goal would never have been reached. Finally, I would like to especially thank Dr. Norma Mertz, committee chair, who from the very beginning of my course of study, continually inspired me to "think about" the evidence and alternatives to my own thinking. She is a true critical thinker.

There have been many others who have helped in a significant way. Gratitude is owed to Dr. Karen Adsit who provided technical support for this project; to Dr. Kay Chitty who encouraged me to "take just one course and see"; to many friends and colleagues who assured me that the goal was attainable; to Dr. Cherry Guinn who allowed me to use the data set which she had compiled and who supported me along the way. I would also like to thank the other doctoral students who shared in this journey: Judy Fry, Dr. Dave Tilley, Dr. Ruth Liu, Dr. Janet Secrest, and Dr. Virginia Keatley.

Lastly, I am very grateful to my husband, Steve, who never complained when there was no room on the table to eat because of all the books and papers and when there was no one to go with him to share in his hobbies.

ABSTRACT

Critical thinking, an engagement in thought requiring examining, questioning, and reflecting on ideas, has become a prominent theme at all levels of education. While there is little agreement about its precise meaning, there is considerable agreement that the concept is desirable, particularly in higher education. In the 1980s prominent educators as well as several national groups argued for an improvement in the critical thinking abilities of college students. Nursing educators and organizations have also strongly advocated that graduates of nursing programs be able to think critically.

Even without a clearly accepted definition, there has been considerable research on the effect of nursing curricula on students' abilities to think critically. The instrument most frequently used in these studies has been the Watson Glaser Critical Thinking Appraisal (WGCTA). The findings from this research have been mixed, leading to conflicting and inconclusive findings. Accounting for these conflicting findings is an important area to explore.

Since the majority of studies have used the WGCTA, the conflicting findings may be due to the limitations of the designs employed to conduct the research. Thus, the purpose of this study was to compare the effectiveness of a longitudinal repeated measures design model with a pretest/posttest design model in assessing changes in critical thinking abilities of a cohort of baccalaureate nursing students. The research questions guiding this study were: (1) does a pretest/posttest design lead to the same conclusions about changes in total critical thinking scores as a repeated measures design? and (2) does a pretest/posttest design lead to the same conclusions about changes in component critical thinking scores as a repeated measures design?

An existing data set, consisting of WGCTA scores obtained from a cohort of baccalaureate nursing students from 1992 through 1994 who were enrolled in a public,

comprehensive university in the southeast, was used to answer the research questions. This data set allowed for the comparison of a pretest/posttest design with a repeated measures design. With alpha set at 0.05, paired t-tests were used to evaluate the pretest/posttest design and repeated measures analysis of variance was used to evaluate the repeated measures design.

The analysis of the data based on a pretest/posttest design for total WGCTA scores revealed no significant differences, however, when the data were analyzed using the repeated measures design, there were significant differences. The analysis of the subtest scores revealed no significant differences with the pretest/posttest design with the exception of the subtest for recognition of assumptions which showed a significant increase. However, when the subtest scores were analyzed using the repeated measures design, significant differences were found for all the subtests with the exception of inferences, which showed no significant difference. Therefore, the answers to both of the research questions are no, the pretest/posttest design does not lead to the same conclusions as a repeated measures design for either the total WGCTA scores or the subtest scores.

The conclusion drawn from this study is that a repeated measures design provides a better method for examining critical thinking than does a simple pretest/posttest design if faculty are truly interested in determining the effect that the nursing curriculum has on the development of critical thinking abilities of students. The repeated measures design is able to show changes which occur across time that are hidden by a simple pretest/posttest design. With this full view of what is happening to critical thinking abilities as students proceed through a course of study, faculties can make informed decisions about necessary changes which need to be made to foster and facilitate this ability in their students.

TABLE OF CONTENTS

CHAPTER	PAGE
I. INTRODUCTION.....	1
Statement of the Problem	12
Purpose of the Study	13
Research Questions.....	13
Significance of the Study	13
Delimitation.....	14
Definitions	14
Methods and Procedures.....	15
Organization of the Study.....	16
II. REVIEW OF RELATED LITERATURE.....	17
Importance of Critical Thinking in Undergraduate Higher Education.....	18
Importance of Critical Thinking in Nursing Education.....	20
Definitions of Critical Thinking	22
The Study of Critical Thinking in Nursing	28
III. METHODOLOGY.....	42
Design of the Study	42
Instrument	43
Protection of Human Subjects	46
Description of the Data Set	46
Data Analysis.....	47
IV. ANALYSIS AND FINDINGS	51
Statistical Analysis.....	52
Pretest/posttest Results.....	52
Repeated Measures Results.....	54
V. DISCUSSION, CONCLUSION, AND RECOMMENDATIONS.....	64
Discussion.....	66
Conclusion	70
Implications for Practice	70
Recommendations for Further Research.....	71
REFERENCES.....	73
APPENDICES	84
Appendix A: Permission To Use Data Set	85
Appendix B: IRB Approval.....	86
VITA.....	88

LIST OF FIGURES

FIGURE	PAGE
1. Repeated Measures Analysis of WGCTA Total Scores.....	56
2. Repeated Measures Analysis of WGCTA Inference Scores.....	59
3. Repeated Measures Analysis of WGCTA Assumption Scores.....	59
4. Repeated Measures Analysis of WGCTA Deduction Scores.....	61
5. Repeated Measures Analysis of WGCTA Interpretation Scores	62
6. Repeated Measures Analysis of WGCTA Argument Scores	63

CHAPTER I

INTRODUCTION

Critical thinking, an engagement in thought requiring examining, questioning, and reflecting on ideas, has become a prominent theme at all levels of education since the early 1970s. While there is little definitive agreement about precisely what it is, there is considerable agreement that it is desirable, particularly in higher education. In the 1980s, several national groups argued for an improvement in the critical thinking abilities of college students. The Rockefeller Commission on the Humanities (1980) called for educational policymakers at all levels to require critical thinking as a basic skill for all students. The National Institute of Education (1984), in explaining the educational outcomes which colleges and universities should strive to achieve, asserted that students need to be prepared to function in a constantly changing world. They went on to explain that critical thinking, the ability to synthesize vast amounts of information, and the mastery of language skills, were the abilities which would allow students to achieve this flexibility. Further, they stated that these skills should be clearly defined, articulated, and measured by the faculty. The Association of American Colleges, in their report entitled Integrity in the College Curriculum: A Report to the Academic Community (1985), argued that all students should learn "...to reason well, to recognize when reason and evidence are not enough, to discover the legitimacy of intuition, (and) to subject data to the probing analysis of the mind" (p. 15).

Prominent educators have added their voices regarding the desirability of critical thinking. Ernest Boyer (1987) of the Carnegie Foundation for the Advancement of Teaching, in reporting the findings and recommendations from his study of the way in which the structures and procedures of college affect the lives of baccalaureate students,

stated that critical thinking along with "clear and effective writing...are...the most essential skills both for further education and for work" (p. 32). In discussing the outcomes of the undergraduate experience, he reasoned that the language skills that the student must possess should be of the higher order and include "...the capacity to think critically, draw inferences, and convey, through effective written and oral communication, subtle shades of meaning" (p. 257). Boyer further stated that if critical thinking is not learned that "...we run the risk of making critical decisions, not on the basis of what we know, but on the basis of blind faith in one or another set of professed experts" (p. 280).

Nursing educators and organizations have strongly advocated that graduates of nursing programs be able to think critically. Carol Lindeman, past president of the National League for Nursing (NLN), has suggested that the ability to think critically is one of five emerging patterns which will have increased importance in the rapidly changing health care system and one which must receive increased attention in schools and colleges of nursing (1995). Kaktaoka-Yahiro and Saylor (1994) have argued persuasively that nurses need to be able to think critically "...in order to be safe, competent, skillful practitioners in their profession" (p. 351). Heaslip (1994) reasoned that nurses who can think critically are able to reason about nursing practice and raise challenges about nursing and the health care system at large in order to provide for client survival and advancement of the profession as a whole.

In 1983 the National League for Nursing (NLN), to which colleges and schools of nursing submit voluntarily for accreditation, required baccalaureate nursing curricula to reflect teaching strategies which promote critical thinking and decision making. In 1989 the NLN further advocated critical thinking as a goal of nursing education by requiring that evidence of curriculum content as well as teaching strategies related to critical thinking be provided. Finally, in 1992, the NLN moved from simply requiring

that the curriculum reflect content and teaching strategies for critical thinking to requiring that nursing programs must not only provide a definition of critical thinking appropriate for their particular program, but that they must also “(p)rovide a rationale and assessment of the methods or processes used to evaluate or measure critical thinking (and that they) report critical thinking outcome data and its use in the development, maintenance and revision of program/s” (p. 26).

Clearly, educators and national groups in both nursing and education are in agreement that critical thinking ability is a necessary outcome of education. And this despite the fact that there is no agreement on a definition of what it is.

Since the early 1950s, definitions of critical thinking have been proposed and debated (Dressel and Mayhew, 1954; Watson and Glaser, 1964 and 1980; McPeck, 1981; Ennis, 1985; Siegel, 1988; Facione, 1990; Pascarella and Terenzini, 1991; Paul, 1993). Nursing educators have similarly attempted to define critical thinking (Bandman and Bandman, 1988; Alfaro-LeFevre, 1995). Despite these efforts, the nursing profession, like other disciplines, has not achieved agreement on a definition.

Nevertheless, even without a clearly accepted definition, there has been considerable research on the effect of nursing curricula on students’ abilities to think critically. The definition used most often by nursing researchers is the one developed by Watson and Glaser (1980) on which the Watson Glaser Critical Thinking Appraisal (WGCTA) is based. Their definition stated that critical thinking is a composite of attitudes, knowledge, and skills. The attitude referred to is described as one of inquiry that involves the ability to recognize problems and to accept the need for supporting evidence regarding that which is asserted to be true. The knowledge required includes knowledge of the nature of valid inferences, abstractions, and generalizations where the accuracy or the weight of the evidence is logically determined. The skills which are needed are those used in applying these attitudes and knowledge. Since many nursing

education programs identify critical thinking in a problem solving context using an adaptation of the scientific method called the nursing process, the use of the definition developed by Watson and Glaser as an underlying framework to evaluate students' abilities to think critically has been seen as appropriate (Jones & Brown, 1991).

The research related to critical thinking in nursing to date has centered primarily around four categories: (1) relationships between critical thinking and various other variables (Tiessen, 1983; Waite, 1989; Miller, 1992; Behrens, 1996; (2) relationships between critical thinking and other professional characteristics in both nursing students and practicing nurses (Matthews & Gaul, 1979; Grossman, 1985; Brooks & Shepherd, 1990; Maynard, 1996); (3) comparisons of critical thinking abilities of students from different types of nursing programs (Frederickson & Mayer, 1977; Pardue, 1987; Brooks & Shepherd, 1990; Brooks & Shepherd, 1992); and (4) effects of nursing education on critical thinking ability (Berger, 1984; Gross, Takazawa, & Rose, 1987; Bauwens & Gerhard, 1987; Sullivan, 1987). This research has produced inconclusive and conflicting results not only about the success of nursing curricula in teaching or enhancing critical thinking, but also about the relationship between critical thinking and other variables deemed important to the nursing profession.

These conflicting results are perhaps most evident in studies which look at the effect of nursing education on critical thinking abilities, all of which used the WGCTA. Berger (1984), using a pretest/post-test design and the WGCTA to compare the critical thinking abilities of sophomore baccalaureate nursing students (BSN) students to those of liberal arts freshman and seniors, found: (1) that sophomore BSN students scored significantly higher than either freshman or senior liberal arts students (however, the method of data collection for the liberal arts students was not specified), (2) that there was a significant increase (measured as gain scores) in critical thinking scores of the

BSN students from the sophomore year till their senior year, and (3) that there was no relationship between critical thinking scores and grade point average (GPA). Likewise, Gross, Takazawa, and Rose (1987), using the same design and instrument, found significant differences on WGCTA entry and exit scores of both technical or associate degree nursing (ADN) students and professional or baccalaureate (BSN) students. They claimed that these findings supported the position that nursing education had a positive influence on critical thinking abilities.

However, other studies have produced contrary findings. Sullivan (1987), using the same pretest and posttest design and the WGCTA instrument, found no difference in critical thinking scores of registered nurse (RN) students in a baccalaureate nursing program between entry and exit. Fleege (1987), using a cross-sectional design and the WGCTA to compare the critical thinking scores of sophomore, junior, and senior nursing students in a five year BSN program, found no significant differences among the scores of the groups. Bauwens and Gerhard (1987), using a pretest/posttest design and the WGCTA instrument, also found no significant differences between the WGCTA entry and exit scores of students in a baccalaureate nursing program. More recently, Maynard (1996), using a cross-sectional design and the WGCTA instrument, found no significant differences in WGCTA scores of baccalaureate nursing students between the time of their entry into the program until graduation.

Research in the other three categories has also yielded similarly conflicting results even though the WGCTA was the instrument utilized in all the research. In studies exploring the relationship between critical thinking and various other variables, Waite (1989) used a one-shot posttest design and found that GPA was positively correlated with the critical thinking abilities of senior baccalaureate nursing students as measured by the WGCTA. A study conducted by Behrens (1996) using a pretest/posttest design and the WGCTA instrument supported the findings of Waite.

The relationship between Scholastic Aptitude Test (SAT) verbal scores, SAT quantitative scores, GPA, age, total number of credit hours in the natural, behavioral and social sciences, the arts and humanities, and the number of credit hours in professional nursing courses was explored by Tiessen (1987) using a one-shot posttest design and the WGCTA instrument. She found that SAT quantitative score, the number of hours in the arts and humanities, and GPA contributed most strongly to the critical thinking abilities of baccalaureate nursing students. On the other hand, Miller (1992), using the same instrument and a pretest/posttest design, found no relationship between GPA and WGCTA scores in registered nurse students enrolled in a baccalaureate program.

Several researchers, using the WGCTA, have investigated the relationship between critical thinking abilities and other professional characteristics, the third category of nursing studies. Brooks and Shepherd (1990), using a one shot posttest design, found a low positive correlation between critical thinking and clinical decision making in students from all types of nursing programs as well as a significant relationship between critical thinking and professionalism. Additionally, Grossman (1985), using a one-shot posttest design, found high critical thinking scores to be significantly related to high problem solving scores. Conversely, Gunning (1981), using the same type of design, found no relationship between critical thinking and clinical problem solving. In order to explore the relationship between critical thinking and moral judgment, Ketefian (1981), utilizing the same type of study design, found critical thinking to be highly correlated with moral judgment while Fleegeer (1987), in contrast, using a cross-sectional design, found only a weak relationship between the two. These studies, like others, revealed conflicting results.

In comparing students from different types of nursing programs, Scolavento (1981), using a one-shot posttest design, found that baccalaureate nursing students'

critical thinking scores were significantly higher than were those of associate degree or diploma students. Conversely, Jones (1985), using the same design, found no difference between baccalaureate and associate degree students critical thinking scores. Both of these researchers used the WGCTA instrument to measure critical thinking abilities.

It is evident that the research surrounding critical thinking and nursing has led to conflicting and inconclusive results. What accounts for the conflicting findings among these studies? Since the same instrument, the WGCTA, was used to measure critical thinking in all the studies, the conflicting results cannot be attributed to the use of differing instruments which could have been measuring different constructs. On the other hand, all of these studies used one of three primary methodological designs, cross-sectional, pretest/posttest, or one-shot posttest. By their very nature, each of these designs is flawed as a way of studying the development of critical thinking, which might account for the conflicting results.

The designs which have been used by nursing researchers to study critical thinking have clear inadequacies and weaknesses. Cross-sectional designs, one of the most frequently used designs to measure critical thinking, are defined as the collection of data on the variable in question at one point in time from subjects at different levels (Aaronson & Kingry, 1988). The advantages of this type of design are that it eliminates the effects of history as well as test-retest, since only one observation of the dependent variable is taken (Franchak, Franken, Subisak, Kapes, & Sheldon, 1980), and that it is cost and time effective (Aaronson & Kingry, 1988). Nevertheless this design is inadequate for studying the question. The primary weakness in using this type of design when attempting to measure change in a variable is that it is impossible to know what the initial level of the dependent variable was in each group which may result in the comparison of noncomparable groups. Thus one is unable to make valid inferences

about changes in a variable across time due to sampling variability. The greater potential for this random variability "...renders the findings from those data more suspect" (Aaronson & Kingry, 1988, p. 189). This type of design "leaves open the question of whether findings are biased by or are completely an artifact of the design employed" (Aaronson & Kingry, 1988, p. 187-188). Since the reported findings could have occurred merely by chance, there is an increased probability of a Type I error meaning that the researcher could conclude that a difference or a relationship existed when, in fact, the difference or relationship was due merely to chance rather than a true treatment effect (Aaronson & Kingry, 1988).

The second most frequently used design employed to measure critical thinking is a pretest-posttest design where a group of subjects is evaluated pre-treatment (pretest), the intervening variable is presented, and then the subjects are evaluated post-treatment (posttest) (Leedy, 1989). The group of subjects is compared with itself, at two points in time, and the difference between the pretest and posttest scores is examined for change (Kerlinger, 1973). The primary advantage to this type of design is that the threat of selection is eliminated. There is no need to try to match the pretest and posttest groups on every relevant variable since both groups are composed of the same subjects (Mitchell & Jolley, 1988). The problem encountered by the researcher who utilizes this design is that any number of extraneous variables could account for an observed change including history, maturation, and test-retest (Levin, Marascuilo, & Hubert, 1978) and because verifying that the treatment actually accounted for the difference in the two scores is extremely difficult (Freed, 1991).

Campbell and Stanley (1963) argued that many potentially change-producing events (history) may occur between the pretest and the posttest which could account for the change in the dependent variable. History becomes a more plausible explanation for the observed results the longer the time lapse between the pre- and posttest. Maturation,

a second possible intervening variable, refers to the biological and psychological processes naturally occurring within the subjects which vary with the passage of time (Campbell & Stanley, 1963). These are changes which could have occurred without the introduction of the dependent variable, thus the data obtained may be based on maturation rather than the dependent variable in question. Again, the longer the time lapse between the pretest and the posttest, the greater the possibility that this variable is an unwanted source of the variance. A third variable which could account for the observed change is the effect of testing or practice. Here the effect is due to the pretest itself. Students will typically perform better on the second test, even when an alternate form of the test is used without any instruction or other interventions (Campbell & Stanley, 1963).

Two other intervening variables have been recognized as having an effect on this type of design. One identified by Campbell and Stanley (1963) is statistical regression where the observed scores tend to gravitate toward the group mean. Thus students who scored high the first time will tend to score lower the second time and those who scored low will tend to score higher. Kerlinger (1973) explained this variable by stating that the high pretest scores are actually higher than they should be simply due to chance, as are some of the low scores. It is unlikely that these scores would be maintained since their value was due primarily to chance and chance was uncorrelated on the pretest and posttest. Mitchell and Jolley (1988) concluded that “(r)egression toward the mean is such an excellent impersonator of a treatment effect that regression fools most of the people, most of the time” (p. 102). Mortality, the idea that all of the subjects that are pretested are not posttested, can also have an effect on the validity of this type of design (Mitchell & Jolley, 1988). The fact that many of the subjects are not present for the posttest may result in comparing two truly different groups. Kerlinger (1973) explained that it is not simply the fact that these extraneous variables can operate, but that we have

no way of knowing whether or not they have operated and affected the dependent variable because there is no way to control for or test for these influences. Since the influence of these intervening variables on the dependent variable cannot be accounted for, the conclusions which are reached may be invalid. Freed (1991) concurred with Kerlinger, stating that with this type of design the impact of threats to internal validity could not be easily examined or accounted for and could thus lead to conclusions which were not true. Leedy (1985) summed up the usefulness of this type of design by stating that it is "a naive approach which relies upon supposition and coincidence in validating the conclusion" (p. 214).

Other studies have used a one shot posttest design. This design has been identified by Cook and Campbell (1979) as a one-group posttest-only design in an effort to distinguish it from a true case study in which "...many variables are measured at the posttest, contextual knowledge is already rich...and intelligent presumptions can be made about what this group would have been like without X" (p. 96). They contended that without a pretest one cannot conclude that the change was related to the treatment. Additionally, without a control group, the threats to internal validity could not be conceptualized or measured. Campbell and Stanley (1963) have even gone so far as to say that "(i)t seems well-nigh unethical at the present time to allow, as theses or dissertations in education, case studies of this nature" (p. 7). Leedy (1985) has described this design as "...the simplest and most naive ..." type of design which is the "...method behind many superstitious folklore beliefs" (p. 213). He stated that in this type of design, a fact is observed, then the consequence is observed, then the two are linked together as cause and effect.

Based on the evaluation of the designs used in the previous studies of critical thinking in nursing students and the flaws that have been identified with these designs, a

plausible explanation for the conflicting results may lie in the methodological design of the extant studies. Light, Singer, and Willett (1990) contended that

the central problem with studying student development stems from poor research design. Too many researchers try to examine development using data collected at only two timepoints, a problem because the accurate depiction of development requires data collected at three or more timepoints (p. 146).

Since the intent of much the previous research in nursing has been to demonstrate the effect that nursing education has on critical thinking abilities of students, and since this effect is one which is expected to occur over the course of the program of study, it would seem that the most appropriate method to measure this effect is a longitudinal design with repeated measures rather than the designs that have been used.

Longitudinal designs have been defined as ones in which data are obtained at different points in time in order to measure changes over time (Fraenkel & Wallen, 1996). This type of data collection allows the researcher to observe the growth trajectory of the variable in question and to answer the question of not only where the student went from point A to point B but also how they got there (Light, Singer, & Willett, 1990). Ewell (1988) has argued that while useful information can be gained from cross-sectional studies, more studies should be focused on the causal effects of college, and that "...longitudinal studies are still required to unambiguously specify the nature and sequence of different types of college impact" (p. 93).

The use of repeated measures design is acknowledged as a method for reducing error variability, that is, the random variability which is due to the differences between people (Keselman & Keselman, 1993). The dependent variable is measured relative to the average response of a subject on all dependent variables, in effect, allowing subjects to serve as their own control. This allows for increased statistical power which is the ability to detect significant differences which genuinely exist or, put another way, the ability to avoid Type II errors. Thus the power to detect true within-subjects treatment

effects is increased as compared with testing these effects in a between-subjects design (Keselman & Keselman, 1993).

Baltes and Nesselroade (1979) have stated the case for the use of repeated measures designs even more forcefully. They have argued that since change extends through time, the only way to measure intra-individual change is to observe the same individual repeatedly, because one observation does "...not provide the data base necessary for development of the representations of the phenomena of interest" (p. 23). They go on to state that the only other appropriate method to measure this type of phenomenon would be to have a sample of individuals who are perfectly identical which is, of course, impossible when observing groups of people including students.

Even though there have been conflicting findings from studies which have examined the effect of nursing curriculum on critical thinking abilities of students, leaders in both education and nursing, as well as professional nursing organizations and accrediting bodies, continue to mandate that graduates possess this ability and that measures which demonstrate this program outcome be employed. Since use of measurement designs which have clear flaws can lead nursing educators to draw invalid conclusions about the ability of students to accomplish this outcome or about the effectiveness of their programs in achieving this outcome, it is imperative to explore designs which may be more appropriate. If the problem lies in the design employed by the researchers, then the conflict among findings may be resolved by the use of an appropriate design.

Statement of the Problem

Educators across all disciplines in higher education support a need for graduates to be able to think critically and to learn that ability as they progress through a program of study. Nursing educators argue that the ability of nursing students to think critically is not only desirable, but is also a factor in patient survival. Accrediting bodies in

nursing require the measurement of critical thinking abilities of graduates as an outcome of nursing programs. Research on the achievement of this outcome has produced conflicting findings which hinder researchers' abilities to answer the question of what effect nursing curricula have on this outcome.

One possible cause for the conflicting results may lie in the limitations of the designs used to conduct the studies. Therefore, a study comparing the most frequently used design (pretest/posttest), with a longitudinal repeated measures design was done in order to attempt to explain this conflict.

Purpose of the Study

The purpose of this study was to compare the effectiveness of a longitudinal repeated measures design model to a pretest/posttest design model in assessing changes in critical thinking abilities of a cohort of baccalaureate nursing students.

Research Questions

The research questions guiding this study are:

1. Does a pretest/posttest design lead to the same conclusions about changes in total WGCTA critical thinking scores as a repeated measures design?
2. Does a pretest/posttest design lead to the same conclusions about changes in component WGCTA critical thinking scores as a repeated measures design?

Significance of the Study

Earlier studies which have attempted to measure the effect of nursing curricula on the critical thinking abilities of nursing students have resulted in conflicting and inconclusive results. Nevertheless, professional organizations and nursing educators continue not only to value this ability, but to desire to clearly demonstrate that their graduates possess this ability. Thus it is important to discover the reason(s) for the conflicting results and to answer the question of whether or not nursing curricula have an impact on the critical thinking abilities of graduates. If the source of the conflicting

results can be identified, nursing educators will be able to move from arguing whether or not nursing curricula improve critical thinking abilities to determining the level of critical thinking they want their students to attain, discovering the pedagogy which will best help students arrive at this level, and appropriately evaluating the students' level of attainment of such skills. In looking at the design methods used in such research, this study will answer the question of whether or not the source of the conflicting results resides in that design.

By using a longitudinal repeated measures design as a part of the study, we may be better able to determine if there are patterns of change in the development of critical thinking ability, and if nursing curricula have a differential effect on some of the component critical thinking abilities --positive on some and at the same time negative on others, which could result in a seemingly negligible total effect on critical thinking. The aforementioned finding could lead researchers to invalidly conclude either that the curriculum had no effect when, in fact, there was an effect which could not be determined based solely on total scores or that there was an effect which was, in actuality, simply an artifact of the design employed.

Delimitation

The study was confined to an analysis of an extant data set collected from a cohort of nursing majors admitted to a baccalaureate nursing program in the southeast United States, between 1992 and 1994.

Definitions

1. ADN student--associate degree student who is pursuing a two-year technical degree in nursing.
2. BSN student--student who is pursuing a four-year baccalaureate or professional degree in nursing as their initial nursing education program.

3. Cohort--a group of people having some factor in common.
4. Diploma student--student pursuing a certificate in nursing from a hospital-based program in nursing.
5. Nursing process--a method of problem-solving used by nurses which includes assessing or data gathering, analyzing data, goal setting and planning of nursing care, implementing nursing care, and evaluating goals.
6. Registered nurse student (RN completion)--student who has either a diploma or an associate degree in nursing and is returning for a baccalaureate degree in nursing.

Methods and Procedures

The study was comparative in design, allowing one to compare the results of using different methodologies (repeated measures and pretest/posttest) with the same data set. The subjects were a cohort of students ($n=63$) who progressed through a baccalaureate nursing curriculum from 1992 through 1994 in a public university in the southeast United States. The group was given the Watson Glaser Critical Thinking Appraisal, an instrument widely used by nursing researchers to measure critical thinking, at the beginning and at the end of their first course in the nursing major and at the end of each semester until graduation. Seven repeated measures for each student in the cohort were obtained over a two and a half year period resulting in a data set consisting of a pretest total critical thinking score and pretest subtest scores for each student and ending with a posttest total critical thinking score and posttest subtest scores for each student. Alternating, but psychometrically equivalent, forms of the WGCTA were used for each test administration.

Using the SPSS statistical package, the data set was analyzed using both a pre/posttest and a longitudinal repeated measures design method. Pretest and posttest total scores and subtest scores were calculated for the group. Total group scores and

subtest scores were then calculated for each semester. Paired t-tests were used to analyze the data based on a pretest/posttest design to see if significant differences existed. A repeated measures ANOVA was used to analyze the data based on a longitudinal repeated measures design. The results of these statistical tests were then compared.

Organization of the Study

The study is organized and presented in five chapters. Chapter one includes an introduction, statement of the problem, purpose of the study, research questions, significance of the study, a description of the methods and procedures used, delimitations, and definitions. In chapter two a critical review of the literature is presented. Chapter three provides an explanation of the methods and procedures used in the study. The findings of the study are presented in chapter. Chapter five includes a summary of the study and the findings, a discussion of the findings, conclusions, and recommendations for further study.

CHAPTER II

REVIEW OF RELATED LITERATURE

Few topics have engendered as much interest among educators as critical thinking. This fact is evidenced by the large number of works published concerning this concept. Since the 1950s it has captured the attention and imagination of philosophers and educators which has led to an increasing emphasis on developing students' critical thinking abilities at all levels of education. This chapter will provide a critical review of the literature related to critical thinking. It is organized into three parts: (1) the importance of critical thinking in undergraduate education and nursing; (2) attempts to define critical thinking; and (3) the study of critical thinking in nursing.

Scholars have been discussing and publishing information and opinions about critical thinking since Dressel and Mayhew's landmark piece General Education-Explorations in Evaluation (1954). This work was an outgrowth of the efforts of a large number of classroom teachers from nineteen colleges and universities whose goals were to explore and make recommendations about the evaluation and testing needs in higher education, especially in the area of general education. The group was divided into six committees, one of which was the Committee on Critical Thinking. An assumption made by this committee, which proved very quickly to be false, was that the members all had the same definition of critical thinking. After much debate and discussion, they decided to define critical thinking as "...a point of view toward problems and their solutions and a way of thinking about basic problems faced by mankind" (p. 273). They subsequently developed a list of behaviors which they believed described this type of thinking. These behaviors were the ability to define a problem, select information which had bearing on the problem and evaluate the reliability of that information,

develop a critical awareness of stated and unstated assumptions surrounding the problem, generate and select appropriate and promising solutions, and draw valid conclusions and inferences.

Based on these behaviors, they developed “A Test of Critical Thinking” which was administered to numerous students in various colleges and universities. They concluded that the test provided a way to study the concept of critical thinking behaviors. Following the publication of the work of this group, increasing emphasis focused on the value, definition, and methods of evaluating critical thinking abilities of students at all levels of education.

Importance of Critical Thinking in Undergraduate (Higher) Education

In reasoning about the value and importance of critical thinking, authorities have posed similar arguments. These arguments are mostly centered around the need for problem solving skills as well as independent thinking. Dressel and Mayhew (1954) contended that all citizens need to be able to think critically in order not only to solve personal problems, but also to maintain a democratic form of government in which the citizens are able to identify problems and solutions independently so that they are not obligated to relinquish their decision-making to an authority. Paul (1984) continued this line of reasoning by stating that critical thinking is necessary in order to “...maintain autonomous judgment in the midst of experts” (p.10). McPeck (1990) echoed the reasons given by Dressel and Mayhew when he asserted that critical thinking was necessary in order to produce informed citizens who are capable of making decisions about problems.

Educators have reasoned that critical thinking is essential for academic as well as vocational success. Ernest Boyer (1987) of the Carnegie Foundation identified critical thinking as one of the most important skills for both further education as well as work.

He explained that without critical thinking skills one would be forced to make decisions based on "blind faith" rather than on the basis of what one knows.

Astin (1993) declared that critical thinking should be the primary cognitive outcome for higher education. Others agreed with Astin. Skinner (1976) stated that education has as one of its' primary goals, to teach students to think. Elaborating on this idea, McMillian (1987), affirmed that higher education should teach students "...to think critically, to reason, and to use judgment effectively in decision making" (p. 3).

Moreover, national groups and organizations have issued statements and reports regarding the role of critical thinking in education. The Rockefeller Commission on the Humanities (1980) called for a heavy emphasis to be placed on critical thinking abilities and for educational policymakers at all levels to require critical thinking as a basic skill. Likewise, the College Board, as well as the Task Force on Education for Economic Growth of the Education Commission of the States, emphasized reasoning as a basic competency (1983).

In addition, a call in undergraduate education for an improvement in the area of critical thinking was issued by the National Institute of Education (1984). Recognizing that graduates have an unlimited amount of new information available to them on a daily basis, they contended that critical thinking was necessary for evaluating and synthesizing this vast amount of information. The Association of American Colleges (1985) echoed this reasoning in their report Integrity in the College Curriculum: A Report to the Academic Community, which stated that all students should learn "...to reason well, to recognize when reason and evidence are not enough, to discover the legitimacy of intuition, to subject data to the probing analysis of the mind" (p. 15). Based on these beliefs, institutions such as the California State University system have gone so far as to make the study of critical thinking a requirement for graduation (Baron & Sternberg, 1987).

Taxpayers, along with employers, are calling for accountability in higher education regarding the ability of graduates to think critically. Faculties are being asked to justify their actions and to evaluate their curricula in light of this desired outcome. This objective received national attention with the adoption of the National Education Goals which were placed into statutory law by the Clinton administration in the form of the "Goals 2000: The Educate America Act" (National Education Goals Panel, 1991). Goal 5.5 calls for a substantial increase in the number of college graduates who can "demonstrate an advanced ability to think critically, communicate effectively, and solve problems" (p. 5).

Importance of Critical Thinking in Nursing Education

Nursing educators have likewise been writing about their concern for nurses' ability to think. Harmer (1928) stated that the goal of nursing education was to assist the student to be able to apply the scientific method to work. She went on to state that use of this method allowed one to solve new problems and adjust to new situations, both of which were a "...measure of our intelligence and our success..." (p. 9-10). Stewart (1944) asserted that nurses need to be able to use their specialized knowledge and skills in order to solve new problems as well as to be able "...to deal with unusual and unpredictable situations...(which) requires more than ordinary resourcefulness and initiative as well as sound professional knowledge and judgment" (p. 337-338). McManus (1949) extended this line of thinking when she claimed that professional nursing "requires special and professional knowledge, technical skill, and judgment that result from specific instruction and the exercise of mental faculties" (p. 2). She continued

(o)nly when the nurse has developed an understanding of the scientific facts and principles underlying nursing, the intellectual techniques for solving professional problems, the social techniques for dealing with people, as well as skilled artistry in the performance of her work--and only when these essential abilities articulate and function simultaneously

and with unity of purpose--can her nursing practice be of a professional nature (p. 4).

As early as 1978, Helen Yura, assistant director for the National League for Nursing, in describing graduates of a baccalaureate nursing program, stated that the practice of these graduates must be saturated with intellectual skills including "...critical thinking..." (p. 5), and that they must be able to engage in "...reasoned judgment..." (p. 6). Many other authors have explained the need for nurses to be able to think critically. Bandman and Bandman (1988) stated that nurses need critical thinking abilities in order to examine various points of view, to identify existing differences and possibilities, and, perhaps more importantly, to enhance the validity, reliability and worth of nursing outcomes. White, Beardslee, Peters, and Supples (1990) asserted that this ability was necessary because of the vast amount of data which the nurse must "...decipher, interpret, and act on (or not)" (p. 16). They added that the crisis in the healthcare delivery system also mandated the need for this ability. Kramer (1993) went on to say that critical thinking was necessary because nurses "...are required to synthesize and integrate multiple forms of knowledge to make health-affirming decisions that embody changing values" (p. 406). More broadly, Kataoka-Yahiro and Saylor (1994) maintained that the ability of nurses to think critically was necessary to provide "...safe, competent, (and) skillful (care)" (p. 351). However the most telling argument was given by Bevis and Watson (1989) who explained the need for this concept in nursing by stating that "...the high levels of technology, the severity of illness, and the critical nature of timing in interventions, make the critical-thinking, expert clinician a significant factor in survival" (p. 354). These views are echoed by many other nursing educators (Schank, 1990; Klaassens, 1988; Case, 1994; Miller & Malcolm, 1990; Heaslip, 1994).

One of the first professional nursing groups to discuss the importance of critical thinking in nursing education was the American Association of Colleges of Nursing. In their publication Essentials of college and university education for professional nursing (1986), they stated that due to the divergence and complexity of nursing practice today, it was necessary for nursing education to prepare nurses who can think critically and creatively in order to make informed decisions and ethical choices regarding societal problems, health care, and the nursing profession.

In 1983 the National League for Nursing (NLN), to which colleges and schools of nursing submit voluntarily for accreditation, required that baccalaureate nursing curricula reflect teaching strategies which would promote critical thinking and decision making. The emphasis at that time was on process rather than outcome in that schools were not required to demonstrate whether or not students were able to think critically, but rather only that strategies were utilized to enhance this ability. The NLN further supported critical thinking as a goal of nursing education in 1987 when the Division of Baccalaureate and Higher Degree Programs published a document describing the characteristics of a graduate of a baccalaureate program in nursing which included the ability "...to think critically..." (p. 2). Beginning in 1992 the emphasis began to change from one of process to one of outcome when the League moved the criterion of critical thinking from a previously optional outcome criterion to a required one. This meant that schools and colleges of nursing would have to provide satisfactory documentation of the ability of their students to think critically in order to receive full accreditation (National League for Nursing, 1992).

Definitions of Critical Thinking

Even though there is widespread agreement on the desirability of critical thinking, there has not been similar agreement on the definition. Dressel and Mayhew (1954) were among the first researchers to attempt to define critical thinking. In their

work, they listed a set of abilities which they believed defined critical thinking: identifying and defining central issues or problems, selecting pertinent information for the solution of a problem, recognizing underlying assumptions, formulating and selecting relevant hypotheses, evaluating evidence or authority, and drawing valid conclusions or judging the validity of inferences. They further contended that while critical thinking was positively related to intelligence as well as to content, content knowledge did not guarantee a high level of critical thinking ability.

Following the definition developed by Dressel and Mayhew, many others have proposed and debated definitions. Watson and Glaser (1980), building on the work of Dressel and Mayhew, broadened the definition to include an attitude component which they defined as one of inquiry and involved the ability to recognize a problem and to accept the need for supporting evidence regarding that which was asserted to be true. The two other components included in their definition were knowledge and skills. The necessary knowledge included knowledge of the nature of valid inferences, abstractions, and generalizations where the accuracy or the weight of the evidence was logically determined. They identified the skills involved only as those used in applying these attitudes and knowledge but did not further elaborate. Based on this definition, they developed the Watson Glaser Critical Thinking Appraisal (WGCTA) instrument which has been used extensively by researchers in attempting to measure critical thinking abilities of students.

Due to what he called a “lack of careful attention to the concept, critical thinking” (p. 81), by researchers in both education and psychology, Ennis (1962) developed his initial definition of critical thinking as “...the correct assessing of statements...”(p.83) and included a list of twelve aspects or abilities which he believed were necessary for critical thinking. These aspects were

1. Grasping the meaning of a statement.

2. Judging whether there is ambiguity in a line of reasoning.
3. Judging whether certain statements contradict each other.
4. Judging whether a conclusion follows necessarily.
5. Judging whether a statement is specific enough.
6. Judging whether a statement is actually the application of a certain principle.
7. Judging whether an observation statement is reliable.
8. Judging whether an inductive conclusion is warranted.
9. Judging whether the problem has been identified.
10. Judging whether something is an assumption.
11. Judging whether a definition is adequate.
12. Judging whether a statement made by an alleged authority is acceptable. (p. 84).

This definition made more explicit the skills which had been identified by earlier writers. In 1985, he revised his definition and stated that critical thinking was “reasonable reflective thinking that is focused on deciding what to believe or do” (p. 14). He broadened his interpretation to include thirteen dispositions (in addition to the abilities he had identified previously) which he believed a critical thinker must possess. These dispositions included such things as seeking reasons, using credible sources, open-mindedness, and sensitivity to others. He also believed, like Watson and Glaser, that critical thinking was a transferable skill and not dependent upon knowledge in a specific subject.

In spite of the fact that previous definitions had been developed, McPeck (1981) argued that critical thinking remained a vague concept which needed further definition. Consequently, he developed a definition of critical thinking in which he stated that it was “the appropriate use of reflective skepticism linked with specific areas of expertise and knowledge” (p. 24). This definition was similar to others but varied in one important aspect, namely, the generalizability of the skill. McPeck believed that critical thinking was not a generalizable skill but rather was based on particular activities learned in a specific field of study. He viewed critical thinking as the resolution of a problem which involved knowing how, in addition to knowing that, and argued that it was not simply a matter of asking questions but rather of knowing when to question as well as what to

ask. He later asserted that critical thinking involved an affective domain as well as a cognitive one (McPeck, 1990). The affective domain consisted of "...reflectively appreciat(ing) the strengths and limitations of (one's) own knowledge" (p. 16). He went on to state that the cognitive skills involved in critical thinking could be taught through drills and problem solving exercises in a discipline. However, he contended that the obstacle that educators continued to face was how to get students to "appreciate" or value and thus use critical thinking skills.

Others have attempted to refine or explain previous definitions. Siegel (1988), who defined a critical thinker as one who was "appropriately moved by reasons" and who "has the ability properly to assess the force of reasons in the many contexts in which reasons play a role" (p. 23), agreed with McPeck, in that the critical thinker must possess a critical attitude or a "critical spirit", which he explained was "the willingness, desire, disposition" to base one's beliefs and actions on reason (p. 23). More recently, Pascarella and Terenzini (1991) have stated that critical thinking involves the ability to "...do some or all of the following: identify central issues and assumptions in an argument, recognize important relationships, make correct inferences from data, deduce conclusions from information or data given, and evaluate evidence or authority" (p. 118). Richard Paul (1993) has postulated a somewhat more convoluted definition by stating that critical thinking is "thinking about your thinking while you're thinking in order to make your thinking better" (p. 91).

Due to the fact there was not a stated agreement on one definition of critical thinking, an attempt was made to arrive at a consensus about the definition by the American Philosophical Association (Facione, 1990). Through the use of the Delphi technique, a panel of 46 scholars, educators, and leading figures in critical thinking theory and research selected from throughout the United States and Canada, worked for two years to develop the definition. They agreed that critical thinking consisted of both

cognitive skills and affective dispositions. The cognitive skills include “...(1) interpretation, (2) analysis, (3) evaluation, (4) inference, (5) explanation, and (6) self-regulation” (p. 9). The affective dispositions identified were:

inquisitiveness with regard to a wide range of issues,
 concern to become and remain generally well-informed,
 alertness to opportunities to cue critical thinking,
 trust in the processes of reasoned inquiry,
 self-confidence in one’s own ability to reason,
 open-mindedness regarding divergent world views,
 flexibility in considering alternatives and opinions,
 understanding of the opinions of other people,
 fair-mindedness in appraising reasoning,
 honesty in facing one’s own biases, prejudices, stereotypes,
 egocentric or sociocentric tendencies,
 prudence in suspending, making or altering judgments,
 willingness to reconsider and revise views where honest
 reflection suggests that change is warranted,
 clarity in stating the question or concern,
 orderliness in working with complexity,
 diligence in seeking relevant information,
 reasonableness in selecting and applying criteria,
 care in focusing attention on the concern at hand,
 persistence though difficulties are encountered,
 precision to the degree permitted by subject and
 circumstances (p. 25-26).

It is apparent that there is not a single definition to which all agree. Nevertheless, all of the above definitions are similar in that they view the outcome of critical thinking as either a decision or the resolution of a problem. They differ on whether this type of thinking requires knowledge in a specific discipline and/or whether it requires a willingness to exercise these abilities.

Similarly, nursing educators have also struggled with creating one definition which is broadly accepted across the discipline. Most of the definitions developed by nurses are based on or adapted from previous definitions. In 1988, Bandman and Bandman defined critical thinking as a “...rational examination of ideas, inferences, assumptions, principles, arguments, conclusions, issues, statements, beliefs, and actions” (p. 5). They further stated that “(t)his examination covers scientific reasoning,

includes the nursing process, decision making, and reasoning in controversial issues” (p. 5). In essence, they defined critical thinking as problem solving and decision making. In this respect their definition was very similar to the one proposed by Watson and Glaser as well as Ennis.

Alfaro-LeFevre (1995), another nursing educator, defined critical thinking as “...purposeful, goal-directed thinking that aims to make judgments based on evidence (fact) rather than conjecture (guesswork)” (p. 9). She agreed with McPeck that the appropriate use of critical thinking skills required knowledge within the discipline. Additionally, she asserted that critical thinking was based on principles of science and scientific method which she equated with maintenance of a questioning attitude and that there were certain attitudes or characteristics which a critical thinker must possess including open-mindedness, curiousness, and flexibility to name a few. These characteristics were similar to the “critical spirit” identified by Siegel. However, in comparing critical thinking to problem solving, she asserted that critical thinking is more open-ended and involved problem prevention as well as the maximization of potential and efficiency.

Jones and Brown (1991) surveyed the deans and directors of National League for Nursing accredited baccalaureate and higher-degree programs in nursing regarding their definition of critical thinking. Their findings, based on a rather low (51%) response rate, revealed that there was no universally accepted definition among nursing educators, but the predominant model of critical thinking in nursing education was the problem solving model. Thus it seems that nursing educators, like others, have not reached agreement on one definition or framework which is domain-specific and incorporates all areas of nursing (Kataoka-Yahire, & Saylor, 1994; Rane-Szostak & Robertson, 1996).

The Study of Critical Thinking in Nursing

In spite of the lack of agreement on a single definition, nursing researchers have continued to study the concept. Since nursing education programs continue to identify critical thinking in a problem solving and decision making context using an adaptation of the scientific method called the nursing process, the definition developed by Watson and Glaser has been the predominant one used as an underlying framework for nursing research in attempting to evaluate students' abilities to think critically.

During the 1980s, in response to increased emphasis on critical thinking by the national accrediting agency, colleges and schools of nursing began to aggressively explore ways in which to measure students' ability to demonstrate this competency and to compare critical thinking abilities to other variables. The studies investigating critical thinking in nursing can be generally divided into four major areas: (1) the relationship between critical thinking and other variables; (2) the relationship between critical thinking and other professional characteristics in both nursing students and practicing nurses; (3) the comparison of critical thinking abilities among students from different types of nursing education programs; (4) the effects of nursing education on critical thinking abilities. While the findings from some of these studies would lead to clear conclusions regarding critical thinking, others point out either modest or clear discrepancies.

The instrument used almost exclusively by nursing researchers has been the Watson Glaser Critical Thinking Appraisal (WGCTA). Watson and Glaser (1980) stated that the instrument may serve any one or more of the following purposes: (1) to measure gains in critical thinking ability, (2) to predict success in certain educational programs, and (3) to determine the relationship of critical thinking abilities to other variables. Further, they contended that the instrument requires no knowledge of any

discipline specific content. There are two equivalent forms labeled "Form A" and "Form B".

The instrument consists of eighty (80) multiple choice questions. These items are of two kinds, "neutral" ones which dealt with content about scientific facts, the weather, or other subjects for which participants would not have strong emotional feelings, and "controversial" ones such as political and social issues which would be expected to engender strong feelings (Watson & Glaser, 1980, p. 2). The reason given by the authors for the inclusion of these two types of items was that strong feelings about a subject tend to affect participants ability to think critically. The questions are based on data which could be encountered on a daily basis at work, in a classroom, or in newspaper or magazine articles. These eighty questions are broken down into five subtests: (1) discriminating the truth or falsity of inferences, (2) recognizing assumptions underlying statements or assertions, (3) deducing whether certain conclusions follow from information given in statements or premises, (4) interpreting evidence to decide if generalizations or conclusions are warranted, and (5) evaluating arguments relative to their strength and relevance to an issue. There are sixteen items in each subtest. A total score, which the authors believe represents an estimate of the students' proficiency in critical thinking, as well as subtest scores are calculated based on the number of correct responses. The level of reading difficulty is designed to be at the ninth-grade level (Watson & Glaser, 1980).

Reliability of an instrument is the accuracy and dependability of that instrument in measuring the variable under study (Kerlinger, 1973). The more error in a measurement, the more unreliable is the measurement. Watson and Glaser (1980) used three methods to estimate the reliability of the WGCTA: (1) internal consistency, (2) stability of scores over time, and (3) correlation between scores of alternate forms of the instrument.

The method used to calculate the internal consistency of the instrument was the split-half procedure which involves scoring two halves of a test separately for each individual and then testing to determine the degree to which the two halves provide the same results by calculating the split-half reliability coefficients (Fraenkel & Wallen, 1996). The Spearman-Brown formula was used to correct for test length. A perfect positive correlation (r) would equal +1.00 with an r of .00 indicating no relationship. The coefficients for the WGCTA ranged from .69 to .85 indicating a fairly high positive correlation and thus reasonable reliability (Watson & Glaser, 1980).

A second method to assess reliability used by Watson and Glaser (1980) was the stability of scores over time. The stability of responses over time was determined by administering the instrument twice to a group of college students ($n=96$) with a three month interval separating the testing events. The scores were then correlated resulting in a correlation coefficient of .73, which revealed reasonable stability (Watson & Glaser, 1980).

Finally, alternate-form reliability was calculated by correlating the responses of students who took both Form A and Form B. The resulting correlation coefficient was 0.75 indicating strong evidence of reliability (Watson & Glaser, 1980).

In addition to determining the reliability of an instrument, validity must also be considered since an instrument which is reliable but not valid is equally unusable. In order to be considered valid, an instrument must measure what it purports to measure (Kerlinger, 1973). Content validity is not only the degree to which an instrument measures what it intends to measure but also the degree to which it contains items which are representative of the content of the property which is being measured. Content validation is based on expert judgment (Kerlinger, 1973). Content validity of the WGCTA is predicated on the acceptance of the definition of critical thinking asserted by the authors since the instrument is composed of subtests which the authors contend are

the skills inherent in critical thinking. Nevertheless, Watson and Glaser have stated that there is not general agreement on the definition of critical thinking. However, they maintained that teachers who desire to improve the critical thinking skills of their students work from a contextual frame of reference and they asserted that the extent to which the WGCTA measures specific objectives from these programs is an indication of its content validity (Watson & Glaser, 1980). While reviews of the instrument have been generally positive, one reviewer questioned the content validity stating that the test items were somewhat narrow in scope and content (Berger, 1985). Still another reviewer stated that the measurement of construct validity had "...not been explored...thoroughly and systematically" (Helmstadter, 1985, p. 1693).

Construct validity, that is the extent to which factors or constructs explain test performance, attempts to validate the theory underlying the construct in question. Methods useful in accomplishing this goal are correlational studies and factor analysis (Kerlinger, 1973). Watson and Glaser (1980) cite studies in which students who were in programs to enhance critical thinking ability performed higher on the WGCTA than did students who were not enrolled in such programs. Additionally, the WGCTA correlates with general measures of intelligence (.41 to .75) including the Scholastic Aptitude Test, the American College Test, and the College Entrance Examination Board while at the same time factor analysis indicated that it was measuring an ability which is distinct from general intelligence (Watson & Glaser, 1980).

Nursing researchers have used the WGCTA to explore the relationship between critical thinking and other variables ranging from the very tangible, such as Grade Point Average (GPA) and success on the National Council Licensing Examination (NCLEX), to the very abstract, such as moral reasoning. Tiessen (1987) and Waite (1989) both found that critical thinking ability, as measured by the WGCTA, was positively related to GPA for BSN students while Behrens (1996) found the same relationship using the

same instrument with diploma program students. These studies would seem to lead to the conclusion that overall academic achievement as measured by GPA is associated with critical thinking. However, Miller's (1992) study of registered nurse students in a baccalaureate nursing program found no relationship between posttest WGCTA scores and total GPA. Moreover, while examining the use of WGCTA scores to predict the success of baccalaureate nursing students on the National Council Licensing Examination, Bauwens and Gerhard (1987) found, through use of regression analysis, that the entry score on the WGCTA and entry GPA accounted for 15% of the variance in the NCLEX scores ($p = .001$). They concluded that WGCTA might be a potential predictor of licensing success.

In exploring the relationship of less tangible characteristics to critical thinking abilities, Ketefian (1981) and Fleeger (1987) considered the relationship of moral reasoning/judgment and critical thinking. Ketefian (1981) designed a study to discover whether or not there were any relationships among critical thinking, type of educational preparation, and moral judgment of practicing nurses. She found that critical thinking as measured by the WGCTA and educational preparation (either ADN or BSN) accounted for 32.9% of the variance in moral judgment as measured by the Rest's Defining Issues Test and that critical thinking and moral judgment of practicing nurses were positively related ($r = .53$). However, since she did not examine the relationship of critical thinking and the type of educational preparation directly, no conclusions could be drawn about the effect of specific type of educational preparation and critical thinking abilities. Fleeger (1987) explored the relationship between aptitude, achievement, critical thinking, and moral reasoning among three levels of nursing students in one program using a cross-sectional comparative design. The findings indicated, among other things, that there was no significant difference among the groups on critical thinking ability or moral reasoning. However, there was a positive though weak relationship ($r = .28$)

between WGCTA scores and moral reasoning thus marginally supporting the findings of Ketefian. However, since there was no control group, there was no control for the internal threats to validity of history and maturation. These findings seem to suggest that critical thinking is a component of moral reasoning.

Further examination of these studies revealed that Tiessen (1987), Waite (1989), and Ketefian (1981) used a one-shot posttest design, that Behrens (1996), Miller (1992), and Bauwens and Gerhard (1987) used a pretest/posttest design and that Fleegeer (1987) used a cross-sectional design. As stated previously, all of these designs have inherent problems. The one-shot designs do not utilize a pretest which makes it impossible to determine if the effect was due to the treatment under study (Cook & Campbell, 1979). Moreover, the pretest/posttest design studies do not account for extraneous variables such as history and maturation which could account for the observed change in the variable in question (Campbell and Stanley, 1963). In studies which use a cross-sectional design, it is impossible to determine the prior level of the variable and thus may result in comparison of noncomparable groups (Aaronson & Kingry, 1988).

The second area of nursing research in critical thinking concerns the relationship between various professional characteristics and critical thinking in both nursing students and practicing nurses. Many of the findings in this area have been conflicting. Matthews and Gaul (1979), utilizing a one-shot posttest design, the WGCTA, and a researcher developed instrument from which the students identified patient problems, which they labeled as nursing diagnoses, found no relationship between critical thinking and the ability to identify nursing diagnoses for either BSN or graduate students in nursing. Likewise, Gunning (1981) found no relationship between critical thinking ability, as measured by the WGCTA, and clinical problem-solving for BSN students utilizing a one-shot posttest design.

On the other hand, other studies have not supported these findings. Grossman (1985), in trying to identify factors which were associated with the ability to problem solve using the nursing process in baccalaureate nursing students, utilizing a one-shot posttest design, discovered that high critical thinking scores were significantly related to higher problem solving scores using the WGCTA and Spark's Nursing Process Utilization Inventory. Brooks and Shepherd (1990) found a weak but significant positive relationship ($r = .249$) between critical thinking and clinical decision-making in students across four types of nursing education programs using the same type of design and instrument. In a more recent study, these same authors explored the relationship between professionalism and critical thinking with the WGCTA and the Health Care Professional Attitude Inventory instruments (Brooks & Shepherd, 1992). They found a positive correlation ($r = .447$) between professionalism and critical thinking across all programs. Maynard (1996) selected a cross-sectional design to study the relationship between critical thinking and professional nursing competence in beginning students, graduating students, and practicing nurses and found no relationship. Danielson (1985) however, did find a small correlation between critical thinking scores on the WGCTA and professional socialization as measured by the Nurse's Professional Orientation Scale for both ADN and BSN students with a one-shot posttest design. Clearly, these studies do not conclusively support a relationship between critical thinking and various professional characteristics.

Once again the results and conclusions from these studies may be called into question due to the type of design employed by the researchers. All of these studies used a one shot posttest except Maynard (1996) who used a cross-sectional design. All the difficulties with a one shot posttest design may be operating in these studies including an undetermined level of the variable prior to the treatment and the inability to determine or control for threats to internal validity (Campbell & Stanley, 1963).

Considerable research has been conducted comparing the critical thinking abilities of students from different types of nursing programs. The argument prompting these comparisons revolves around whether or not baccalaureate programs in nursing were truly better at producing a nursing professional than the others. The argument for a baccalaureate degree was supported by the American Nurses Association in 1965 when they declared that the baccalaureate degree should be the minimum for entry into the practice of professional nursing (American Nurses' Association, 1965). Nevertheless, research conducted to unequivocally support this position has not been produced. Once again, the findings have been inconclusive and conflicting.

In one of the earliest studies comparing students from different types of programs, Frederickson and Mayer (1977), using a posttest only design, reported that baccalaureate students in their last semester scored significantly higher on the WGCTA than did associate degree students who were also in their last semester prior to graduation. Since this was a pilot study with only fourteen students in the sample, no conclusive generalizations could be made. Similarly, Lynch (1988) compared 74 baccalaureate students with 87 associate degree students in a one shot posttest design and found that the baccalaureate students scored significantly higher ($p < .001$) than the associate degree students on the WGCTA, supporting the findings of Frederickson and Mayer. Pardue (1987), who tested critical thinking using the WGCTA and cross-sectional design in a unique study of licensed nurses from associate, diploma, baccalaureate, and master's programs ($N=121$), found that both the baccalaureate and master's prepared nurses had significantly higher scores than either of the other groups ($p = .05$). Brooks and Shepherd (1990), used the WGCTA, the Nursing Performance Simulation Instrument, and a one-shot posttest design, to study the relationship between critical thinking and decision-making in all four types of basic nursing programs. They found significantly higher ($p = .05$) critical thinking scores for the BSN students and the

RN students than for the diploma and ADN students. These findings were again supported by these authors in a later, similar study (Brooks & Shepherd, 1992). Scoloveno (1981) in a one shot posttest, found that baccalaureate nursing students scored significantly higher ($p < .01$) on the WGCTA than did associate degree or diploma students. Danielson (1985), in comparing baccalaureate and associate degree students' scores on the WGCTA using a convenience sample of 89, found that the baccalaureate students scored significantly higher than did the others. Generalizability was limited due to the small sample size. Opposite results were obtained by Jones (1985), who compared 33 baccalaureate students with 49 associate degree students using the same instrument and same design. She found that the two groups did not differ in their critical thinking scores. Once more, the designs adopted by these researchers could lead to invalid conclusions due to the reasons previously stated.

The discrepancies identified in the previously mentioned areas are not nearly as striking as the ones present in the studies in the fourth area of research, namely, the effects of nursing education on critical thinking. The primary study designs utilized by nursing researchers in trying to determine this effect have been either cross-sectional or pretest and posttest; the instrument most often employed to measure critical thinking skills has been the Watson Glaser Critical Thinking Appraisal (WGCTA).

Aaronson and Kingry (1988) have defined cross-sectional designs as ones in which the data have been collected at one point in time from subjects at different levels. This design has been frequently used by nursing researchers to determine if there is a significant difference between nursing students at different points in the curriculum for the purpose of determining the effect of the curriculum on critical thinking. Using this design, several authors have reported no significant differences in critical thinking. Fleeger (1987) compared critical thinking abilities of students in the upper three years of a five year baccalaureate nursing program (41 students at the beginning of their first

clinical course, 23 students in the second level, and 27 others at the beginning and at the end of the third level for a total sample size of 91) and concluded that there were no significant differences. Brigham (1989) compared the critical thinking abilities of freshman, sophomore, junior, and senior baccalaureate nursing students ($n=114$) and found that there was no significant difference ($p = .0628$) among any of these students, leading to the conclusion that the nursing curriculum had no positive effect on this ability. Similar results were reported by Maynard (1996), who used a cross-sectional design to study the relationship between critical thinking and professional nursing competence in beginning students, graduating students, and practicing nurses and found no relationship. Conversely, Kokinda (1989), using the same instrument and design, found that there was a significant difference in critical thinking skills among four levels of baccalaureate nursing students. This researcher, unlike the previous ones, explored the differences in subtest scores on the WGCTA and found that there were significant differences in inference, deduction, and evaluation of arguments, but no significant difference in recognition of assumptions and interpretation. Nevertheless, since it was impossible to know the initial level of critical thinking of each group, all of these studies have a major inherent design flaw: potential comparison of noncomparable groups on the variable in question (Aaronson & Kingry, 1988). Therefore, valid inferences cannot be drawn from these studies.

Although the findings have been inconclusive in studies using a cross-sectional design, there have been even more marked differences in studies using a pretest and posttest design. While using this type of design and the WGCTA, several authors reported that nursing education had a positive effect on critical thinking abilities of students (Berger, 1984; Soefje, 1985; Gross, Takazawa, & Rose, 1987; Miller, 1992; Pepa, Brown, & Alverson, 1997). Berger (1984) studied 137 baccalaureate nursing students comparing critical thinking scores of sophomores and seniors. She reported a

significant difference ($t=3.98$) between mean scores which increased from 77 for the sophomores to 80 for the seniors. Soefje (1985) conducted a similar study comparing registered nurse students enrolled in a baccalaureate program ($n=246$). She found an increase on four out of five WGCTA subtests with a significant increase on the subtest for evaluation of arguments. These findings were supported by Miller (1992) who also found significant gains for registered nurse students ($p=.05$). Gross, Takazawa, and Rose (1987) explored the impact of nursing education on both associate and baccalaureate degree students ($n=108$). Using a paired t-test they found that there was a highly significant ($p<.000$) improvement in mean WGCTA scores ($t=9.05$) for both groups, but these results must be viewed with caution since they reported a 30% attrition rate. Most recently, Pepa, Brown, and Alverson (1997) compared pre and post WGCTA scores of both traditional baccalaureate and accelerated baccalaureate students. The accelerated students were those students who entered the nursing program with either a completed bachelor's degree in a non-nursing major or had a minimum of 44 college credits and completed the nursing curriculum in 22 months compared with 32 months for the traditional students. The traditional group showed a significant ($p=.007$) increase in the scores, but the accelerated group showed an increase which was not significant ($p=.107$). They concluded that while the nursing curricula supported the development of critical thinking for the traditional students, the shortened time the accelerated student spent in the program might not be enough to have an effect on their critical thinking abilities.

Poole (1989) took a somewhat different approach. She investigated the changes in critical thinking of both baccalaureate and associate degree students by subjecting the WGCTA entry and exit subtest scores of these students to an Incomplete Principle Components Analysis, rotated by the Varimax Criterion, which resulted in the five subtest scores being combined into two factor scores. These two factors were labeled as

High Cognitive (from the subtests of interpretation and evaluation of arguments), which the authors viewed as the more creative skills, and Low Cognitive (from the subtests of inference, recognition of assumptions, and deduction), which the authors viewed as the more logical skills. Students in both types of programs showed an increase in High Cognitive skills ($p=.0005$), but the baccalaureate students showed no change in the Low Cognitive skills ($p=.9052$) with the associate degree students having a negative change which was not statistically significant ($p=.3252$). The author concluded that traditional nursing curricula, with the extensive use of the nursing process, favored logical thinking rather than creative thinking.

Despite these results, there have been just as many studies using the same design and instrument which have revealed opposite findings (Currier-Dagrosa, 1986; Sullivan, 1987; Bauwens & Gerhard, 1987; Kintgen-Andrews, 1988; Notarianni, 1991; Behrens 1996). In a study dealing only with the effect of a specific aspect of the baccalaureate curriculum, Currier-Dagrosa (1986), investigated the effect of a summer work study program on the WGCTA scores of 32 students. The students were tested in May and again in September of the same year. No significant differences were found in the scores. Sullivan (1987), after exploring the effect of baccalaureate curriculum on 51 registered nurse students' critical thinking abilities, found, among other things, that the WGCTA scores remained virtually unchanged ($t=.10$) and concluded that either the entry critical thinking skills of these students may already be high enough so that the curriculum had no effect or that other unmeasured variables had some unknown affect. Bauwens and Gerhard (1987), in examining the use of the WGCTA to predict success on the NCLEX, found that WGCTA scores of baccalaureate students did not change significantly, but these findings must be taken with caution since of the 145 volunteers who took the entry WGCTA, only 53 completed the posttest. Kintgen-Andrews (1988), looked at a broader group of students including licensed practical nursing

students, prehealth science students, associate degree students, and baccalaureate nursing students at the beginning and end of one academic year and found that there were nonsignificant gains observed in WGCTA scores for all students except the baccalaureate students who had a nonsignificant decrease.

Notarianni (1991) compared the changes in critical thinking abilities of students over one academic year in both baccalaureate and associate degree programs. She found that there were no significant increases in the posttest scores of students in any year of the baccalaureate or associate degree programs and, in fact, the scores for students in the second year of the baccalaureate programs were significantly lower ($p=.003$) than those of any other year in those programs. She concluded that nursing education did not seem to facilitate gains in critical thinking for either group of students.

Behrens (1996), studied diploma school students to investigate the relationship between WGCTA scores and academic performance as measured by GPA. He discovered that there was no difference in the pretest scores at the beginning of the program and the posttest scores at the end of the program ($p<.10$), but that WGCTA scores and GPA were significantly related ($p<.01$).

According to Campbell and Stanley (1963), Kerlinger (1973), Freed, Ryan, and Hess (1991), and others, the inherent problem with the use of a pretest/posttest design is that a number of extraneous variables may account for the observed change including history, maturation, and test-retest and thus verifying that the treatment accounted for the difference in the scores is extremely difficult. Therefore, the validity of the studies using this type of design is called into question.

Thus it is easy to discern a clear discrepancy in findings among studies attempting to identify the effect of nursing education on critical thinking abilities of students from all types of nursing programs. Once again, the question begs to be answered: what might account for the differences in findings of the extant studies. One

possible answer is that the design chosen for the study has a consequential effect on the conclusions reached. This study will attempt to answer this question by comparing one of the most frequently used designs (pretest/posttest) with a longitudinal repeated measures design to determine if the conclusions reached would be the same for both types of designs.

CHAPTER III

METHODOLOGY

The purpose of this study was to compare the effectiveness of a longitudinal repeated measures design model to a pretest/posttest design model in assessing changes in critical thinking abilities of a cohort of baccalaureate nursing students. This chapter describes the design of the study, the instrument used, the data set, and the data analysis techniques used.

Design of the Study

The majority of previous studies in nursing examining the ability of students to think critically have utilized a pretest/posttest design and have revealed findings which are conflicting. A comparative design was selected for this study in order to determine the effect of the research design chosen on the findings of the study. Permission to use an existing data set from a cohort of baccalaureate nursing students from 1992 through 1994 who were enrolled in a public, metropolitan, comprehensive university in the southeast, was obtained (Appendix A). This data set consisted of Watson Glaser Critical Thinking Appraisal (WGCTA) scores from the students at the beginning and end of the first nursing course and at the end of each subsequent semester. The first set of scores was used as pretest scores and the last set as posttest scores for the pretest/posttest analysis. The scores obtained at the end of the intervening semesters along with the pretest and posttest scores were used as repeated measures data for the repeated measures analysis. Thus this data set allowed for the comparison of both a pretest/posttest design with a repeated measures design.

Instrument

The Watson Glaser Critical Thinking Appraisal, an instrument widely used in the study of critical thinking (e.g. Bauwens & Gerhard, 1987; Behrens, 1996; Brooks & Shepherd, 1990; Gross, Takazawa, & Rose, 1987; Ketefian, 1981; Kokinda, 1989; Lynch, 1988; Notarianni, 1991; Pepa, Brown, & Alverson, 1997), consists of an 80-item test that can be used for a number of purposes identified by the authors, including the measurement of gains in the ability to think critically. The test is composed of the following five subtests:

1. **Inference.** Discriminating among degrees of truth or falsity of inferences drawn from given data.
2. **Recognition of Assumptions.** Recognizing unstated assumptions or presuppositions in given statements or assertions.
3. **Deduction.** Determining whether certain conclusions necessarily follow from information in given statements or premises.
4. **Interpretation.** Weighing evidence and deciding if generalizations or conclusions based on the given data are warranted.
5. **Evaluation of arguments.** Distinguishing between arguments that are strong and relevant and those that are weak or irrelevant to a particular question at issue. (Watson & Glaser, 1980, p.2).

Subtest scores were recorded as the number of correct responses made by the students. Correct responses were identified in the key provided by the publishers of the test. Total scores were obtained by adding the scores obtained on the five subtests as directed in the WGCTA Manual (Watson & Glaser, 1980) with a potential range from zero to eighty.

The reliability of the instrument, that is the ability of the instrument to produce consistent results, was established by the authors by determining the internal consistency, the stability of responses over time (test-retest), and the correlation between

scores on alternate (parallel) forms (Watson & Glaser, 1980). The authors measured internal consistency by calculating the split-half reliability coefficients and then using the Spearman-Brown to provide an estimate of the reliability of the entire test. In this method the test is randomly divided into two parts and students' scores on each part are compared resulting in a correlation coefficient based on half of the test. In order to estimate the reliability for the whole test, the Spearman-Brown formula is used (Ferguson & Takane, 1989). The coefficients which were obtained ranged from 0.69 to 0.85 which indicated a reasonably strong but not high correlation supporting the notion that the instrument is reliable.

The second method used by Watson and Glaser (1980) to assess reliability was the stability of responses over time, also known as the test-retest method. The underlying premise for this method is that if the test is reliable, the students' scores should remain consistent from one time to the next unless something occurs to change the scores. This analysis was accomplished by administering the instrument twice to a group of college students with a three month interval separating the testing events. The scores were then correlated resulting in a correlation coefficient of 0.73, which revealed reasonable stability again supporting the reliability of the tool.

The third method used by Watson and Glaser (1980) was the correlation between scores on alternate (parallel) forms. Two forms of the test, Form A and Form B, were developed by the authors. Alternate-form reliability was calculated by correlating the responses of subjects who took both Form A and Form B. The resulting correlation coefficient was 0.75, which once again is a strong but not high correlation. Based on this finding, the authors concluded that these forms are equivalent and that they may be used as alternate or interchangeable tests (Watson & Glaser, 1980).

In sum, Watson and Glaser used three methods to assess the reliability of their instrument using three conventional means, namely, internal consistency, test-retest, and

parallel forms. The results of these methods seem to indicate that it is a reasonably reliable instrument.

In addition to assessing the reliability of the instrument, the authors also assessed the validity. Validity of an instrument is most basically concerned with the notion that it measures the construct it claims to measure. Watson and Glaser (1980) examined both the content validity and the construct validity.

Content validity is the degree to which the content of the instrument represents the variable being measured (Kerlinger, 1973). Thus, content validity of the WGCTA is predicated on the acceptance of the definition of critical thinking asserted by the authors who state that there has not been general agreement on a definition. If one does not agree with the definition which the authors propose, then one would not agree that the instrument is a valid measure of critical thinking. They go on to state that even though there is not an accepted definition of the concept, they believe that there is a "contextual frame of reference from which teachers must work" if they are attempting to enhance these abilities (Watson & Glaser, 1980, p. 10). Thus, they claim that one indication of the content validity of the instrument is the extent to which it measures a sample of the course objectives developed by these educators. Nevertheless, they do not explain how they went about examining whether or not the instrument measured these objectives.

Construct validity refers to the extent to which certain factors or constructs explain test performance and attempt to validate the theory behind the construct, often by factor analysis and correlation studies (Kerlinger, 1973). Studies were cited by the authors of the WGCTA in which students who experienced courses designed to increase critical thinking skills showed a greater gain in this ability as measured by the WGCTA than did students who did not receive these types of experiences (Watson & Glaser, 1980).

It is evident that the authors attempted to establish both content and construct validity of the instrument. The results of these attempts is that the instrument is considered reasonably valid.

Protection of Human Subjects

Since this study involved a secondary analysis of data, Form A was submitted to the University of Tennessee Institutional Review Board requesting an exemption from IRB review based on category four, which states that research involving existing data recorded by the investigator in a manner in which the subjects could not be identified may be exempt from review. Permission to conduct the study was requested and obtained (Appendix B) from the University Review Board.

Description of the Data Set

The data used for this study consisted of Watson Glaser Critical Thinking Appraisal (WGCTA) scores obtained between 1992 and 1994 from a cohort of baccalaureate nursing students in a public, metropolitan, comprehensive university in the southeast United States. Students were not required to complete the instrument and the scores obtained did not impact the students' grades. The WGCTA had been administered to these students at the beginning of the first semester following admission to the nursing major and at the end of each subsequent semester. Thus, seven repeated measures for each student in the cohort were obtained over a two and a half year period. This resulted in a data set consisting of a pretest total critical thinking scores and pretest subtest scores and ending with a posttest total critical thinking score and posttest subtest scores from sixty-three students.

One of the threats to internal validity is that of testing. When this threat occurs it is due to the fact that the participants were tested with the same instrument more than once and the change which occurred was due to practice and experience with the instrument rather than to the independent variable (Mitchell & Jolley, 1988). In order to

decrease this threat, the psychometrically equivalent forms A and B of the WGCTA had been used alternately for each test administration.

Data Analysis

Since the individual tests in the data set had not been scored, the tests were hand scored by the researcher utilizing the scoring key provided with the WGCTA manual. The subtest scores and the total scores were calculated, coded, and entered directly into the Statistical Package for the Social Sciences 6.1 (SPSS) program for statistical analysis in order to answer the research questions.

Upon examination of the data set, it was found that some data points were missing. These data points were randomly distributed throughout the data set. Various methods have been described for handling missing data including deleting cases, using a missing data correlation matrix, and estimating missing data (Tabachnick & Fidell, 1983). If there are only a few cases with missing data and these cases occur randomly, then it is reasonable to simply drop those cases (Tabachnick & Fidell, 1983). Since there were nine students who did not take any of the tests, either after the second, third, or fourth time the instrument was administered, and these nine cases were randomly scattered throughout the data set, these cases were eliminated from the study. There were four students who did not take two consecutive tests who were likewise omitted from the study. There were seven students who failed to take only one of the tests. Tabachnick & Fidell (1983) state that "(b)ecause, in the absence of all other information, the mean value is a best guess about a missing score on a variable, means are frequently inserted for missing values" (p. 71). Based on this premise, the researcher estimated the missing test values for these seven students by calculating an average or mean of that student's score based on the test score from the test immediately prior and the test score from the test immediately after the test which contained the missing value. This value was then inserted in place of the missing value.

Descriptive statistics were limited to gender and the range, mean, median, and mode of ages of the participants. The mean total scores and mean subtest scores were calculated for each semester. Alpha was set at 0.05. The data set was then analyzed using both a pre/posttest and a longitudinal repeated measures design method.

In order to analyze the pretest/posttest design, a paired or dependent t-test was performed with both the total critical thinking scores as well as the subtest scores. This parametric statistic was chosen to compare the mean scores of the students before and after exposure to the nursing curriculum in order to determine if there was a significant change in those scores as a result of that exposure. This method reflects the standard procedure used in a one-group pretest/posttest design (Freed, Ryan, & Hess, 1991).

A one-way repeated measures analysis of variance (ANOVA) was used to analyze the data based on a longitudinal repeated measures design. This statistic allows for the examination of the dependent variable in terms of the variance attributed to the between-subject variance (variation between the different subjects) and the within-subject variance (variation for the same subject across time) and is the appropriate analysis to use in longitudinal designs when the subjects are measured on multiple occasions (Freed, Ryan, & Hess, 1991).

A post hoc analysis using Tukey's Honestly Significant Difference (HSD) with alpha again set at 0.05 was executed for the ANOVAs which were statistically significant. This procedure was performed to determine which pairs of means were significantly different. This statistic was chosen because it allowed for all possible pairwise comparisons while at the same time, maintaining an acceptable level of the family-wise Type I error rate without an excessive decrease in power (Sheskin, 1997). The family-wise Type I error rate is "the likelihood that there will be at least one Type I error in a set of c comparisons" (Sheskin, 1997, p. 340). All pair-wise comparisons were tested since the researcher did not have a hypothesis that focused attention on a

particular pair of means or combination of means. It must be noted that since this study used a single-factor within-subjects analysis of variance rather than a single-factor between-subjects analysis, the mean square residual was used as the measure of variability. Thus, the appropriate error term to use in calculating the post hoc comparisons was the mean square residual rather than the mean square within-groups (Sheskin, 1997).

The SPSS package does not allow for post hoc comparison procedures with a repeated measures ANOVA. As a consequence of this fact, the researcher entered all the means calculated in the ANOVA's into an Excel spreadsheet program, entered the Tukey's HSD formula to calculate the Studentized Range Statistic, represented by the notation q , and then compared this number with a Table of the Studentized Range Statistic for the 0.05 critical value. The Studentized Range Statistic is the statistic used to test whether or not the differences between the pairs of means are statistically significant. The calculated q value must be equal to or greater than the tabled value of q in order to reject the hypothesis that there is no difference between the means.

In sum, an existing data set composed of WGCTA scores was analyzed. The first set of scores was used as pretest scores while the last set was used as posttest scores. In order to answer the first research question (does a pretest/posttest design lead to the same conclusions about changes in total critical thinking scores as a repeated measures design) a paired t-test was computed using the mean of the first set of total scores compared to the mean of the last set of total scores to determine if there was a significant difference. Next a repeated measures analysis of variance was computed using all seven means for the total scores, in order to determine if there was a significant difference between the means. These results were then compared to determine if the analysis led to the same conclusions. The same procedure was then followed for each WGCTA component score to answer the second research question (does a

pretest/posttest design lead to the same conclusions about changes in component critical thinking scores as a repeated measures design).

CHAPTER IV

ANALYSIS AND FINDINGS

The purpose of this study was to compare the effectiveness of a longitudinal repeated measures design model with a pretest/posttest design model in assessing changes in critical thinking abilities of a cohort of baccalaureate nursing students. The study utilized an existing data base composed of Watson Glaser Critical Thinking Appraisal (WGCTA) scores in order to compare the two types of design models. Chapter IV details the analysis and findings of the study. It begins with a description of the sample and the statistical procedures used, then reports the results in terms of the two types of designs studied.

The data base consisted of Watson Glaser Critical Thinking Appraisal scores for a cohort of baccalaureate nursing students obtained from 1992 through 1994 (n=63). Upon examination of the data set, only fifty (50) cases were valid for inclusion in this study. Since the underlying question was to assess the changes in critical thinking abilities, students who did not complete any tests after either the second, third, or fourth administration were deleted from the study (n = 9), as were students who did not complete two consecutive tests (n=4). Of the remaining fifty students, seven (7) did not complete one of the tests, however, means for these missing scores were computed and then used in the data analysis.

The demographic data available from the data set were limited to gender and age. The majority of the subjects were female (80%). The average age of the subjects was 25.32 years with a range of 20 to 44. The median age was 22.5 with a bimodal age of 21 and 22.

Statistical Analysis

The data set was analyzed with the Statistical Package for the Social Sciences 6.1 (SPSS) program utilizing a paired t-test and a repeated measures analysis of variance (ANOVA). Additionally a post hoc analysis using Tukey's Honestly Significant Difference (HSD) was computed for the ANOVAs which were statistically significant.

Pretest/posttest Results

In order to examine the data based on the pretest/posttest design, the data were analyzed using paired t-tests. The paired t-test was chosen as the analysis technique because it is considered the standard procedure for analysis of a one group pretest/posttest design (Freed, Ryan, & Hess, 1991).

The total WGCTA scores were initially analyzed using a paired t-test with alpha set at 0.05. The first set of scores served as the pretest scores and the last set as the posttest scores. The mean and standard deviation for the pretest scores were 54.58 and 8.919 with a range of 36 to 71. The possible range of scores was from 0 to 80. The mean and standard deviation for the posttest scores were 55.82 and 8.113 respectively, with a range of 40 to 73. While the mean scores did show an increase, the paired t-test revealed that there were no significant differences in the scores from the pretest to the posttest ($t = -1.32$, $df = 49$, $p = 0.194$) thus suggesting that nursing education did not promote significant gains in critical thinking. Table 1 shows the results of this comparison for the total WGCTA scores.

The second aspect of the pretest/posttest design involved examining the five subtest scores using the paired t-test comparison. The possible range for each of these subtest scores was 0 to 16. The subtest for inferences was analyzed first using a paired t-test with the first set of scores serving as the pretest scores and the last set of scores serving as the posttest scores. The mean pretest score for this subtest was 8.66

Table 1.

Paired t-test for Total WGCTA Scores

Variable	Mean	SD	t-value	df	p
Pretest	54.58	8.919	-1.32	49	.194
Posttest	55.82	8.113			

with a standard deviation of 2.327, while the mean for the posttest was 9.14 with a standard deviation of 2.587. The results of the paired t-test revealed that there was no significant change in the scores ($t = -1.11$, $df = 49$, $p = 0.274$).

The same data analysis procedure was used for the subtest for **deduction**. The mean pretest score for this subtest was 10.64 while the posttest mean was 10.88. The standard deviation for the pretest was 2.328 and 2.027 for the posttest. The paired t-test analysis revealed no significant difference in the scores ($t = -0.69$, $df = 49$, $p = 0.491$) which was similar to the results obtained for the subtest for inferences.

The examination of the data for the subtest for **interpretation** disclosed findings which were similar to the first two subtests. The mean for the pretest was 12.00 while the posttest mean was 11.78 with a standard deviation of 1.773 and 1.930 respectively. The paired t-test for this subtest indicated no significant difference in these scores ($t = 0.7$, $df = 49$, $p = 0.488$).

Next, the subtest of evaluation for **arguments** was investigated. This analysis showed the pretest mean to be 12.02 and the posttest mean to be 11.90. The standard deviation for the pretest was 2.272 and 2.866 for the posttest. Again, like the subtest for interpretation, the results of the paired t-test indicated that there was no significant change in these scores ($t = 0.31$, $df = 49$, $p = 0.76$).

Lastly, the same method of analysis was employed for the subtest for **recognition of assumptions**. The mean scores were 11.22 for the pretest and 12.10 for the posttest with a standard deviation of 3.872 and 3.489 for each one. This time the paired t-test revealed a significant increase ($t = -2.08$, $df = 49$, $p = .043^*$) in the pretest and posttest scores suggesting a significant gain in this component score.

In sum, as seen in Table 2, the paired t-tests revealed that the only subtest score which showed a significant change was recognition of assumptions. Of the other four subtests, inference and deduction showed a nonsignificant increase while the subtests for interpretation and arguments showed a nonsignificant decrease.

Repeated Measures Results

The next part of the data analysis involved using all seven of the total WGCTA scores as well as the seven subtest scores and subjecting them to a repeated measures analysis of variance (ANOVA). This analytic technique was chosen because it is considered to be the appropriate technique when subjects are measured on multiple occasions in a longitudinal study (Freed, Ryan, & Hess, 1991).

The repeated measures ANOVA for the WGCTA total scores revealed findings very different from the pretest/posttest design. Whereas the pretest/posttest design did not show any significant differences, the repeated measures revealed significant differences ($F = 2.90$, $df = 6/294$, $p = .009^*$) between the means of the total scores as seen in Table 3. When the repeated measures scores were graphed against the pretest/posttest scores, a pattern of change emerged which showed decreases at the end of semesters one, three, and five. Increases were noted at the ends of semester two, four, and at the posttest. This pattern can be seen in Figure 1.

In order to assess which of these increases and decreases were significantly different from each other, a post hoc analysis was done using Tukey's Honestly

Table 2.

Paired t-test for WGCTA Subtest Scores

Subtest	Pretest Mean	Posttest Mean	t-value	p
Assumptions	11.22	12.10	-2.08	.043*
Inference	8.66	9.14	-1.11	.274
Deductions	10.64	10.88	-0.69	.491
Interpretations	12.00	11.78	0.70	.488
Arguments	12.02	11.90	0.31	.760

*p < .05

Table 3.

Repeated Measures ANOVA for Total WGCTA Scores

Source	SS	df	MS	F	p
Between	353.47	6	58.91	2.90	.009*
Within	5970.25	294	20.31		
Total	6323.72	300			

*p < .05

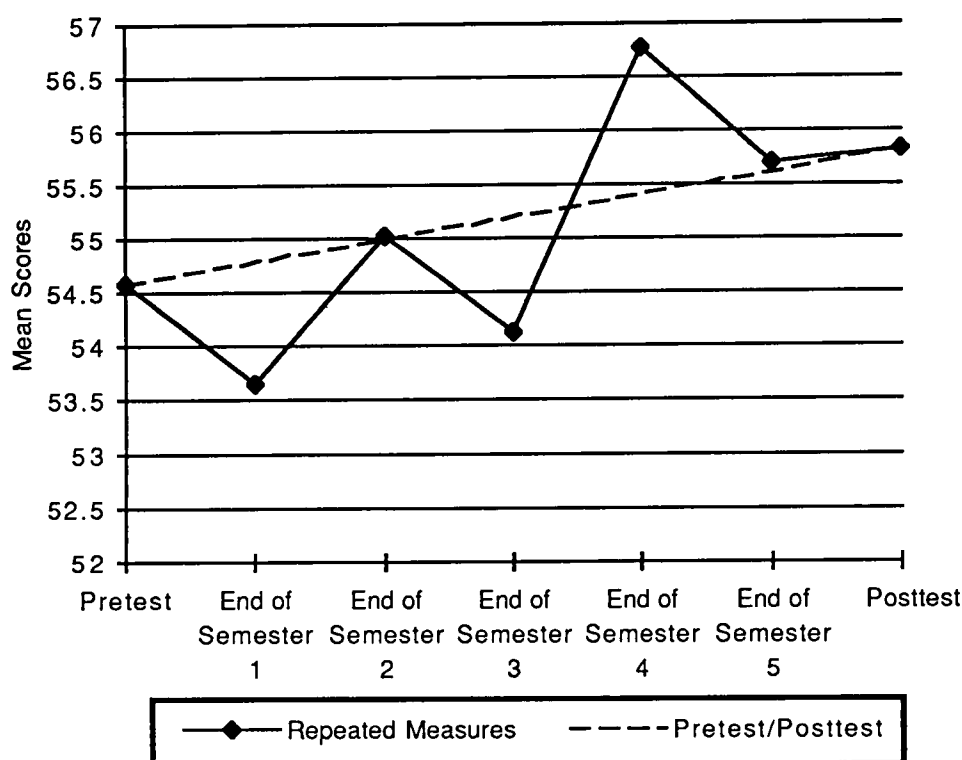


Figure 1. Repeated Measures Analysis of WGCTA Total Scores

Significant Difference (HSD) method. This method was chosen in order to compare all pairs of means to test for significant differences since it had not been hypothesized as to which pairs of means or combination of pairs might be significantly different. This test is the standard procedure to use when an overall significant F had been found in an analysis of variance (Freed, Ryan, & Hess, 1991). The Studentized range or q statistic is used to test if there are any significant differences between pairs of means. The degrees of freedom for the critical value for this statistic is k , which is the number of means, and the number of degrees of freedom within group from the ANOVA (Freed, Ryan, & Hess, 1991). The critical value for the q statistic for this study with $k=7$ and $df>120$ or infinity (since the within group degrees of freedom for this study was 294

and the q table ends with infinity following 120 degrees of freedom), is 4.17 when alpha is set at 0.05. The Tukey's HSD analysis for this study indicated that the students' scores at the end of the fourth semester were significantly different from their scores at the end of the first semester ($q = 4.927^*$) and from their scores at the end of the third semester ($q = 4.174^*$). Therefore, while a pretest/posttest design revealed no significant change in total critical thinking scores over the two and a half year period, the repeated measures design revealed that there were significant differences from one test administration to another.

The second part of the repeated measures analysis looked at the data from the WGCTA subtests. Once again, a repeated measures ANOVA was calculated for each of the subtests using the pretest/posttest scores along with the five intervening scores. The results of this analysis is summarized in Table 4.

The repeated measures analysis for the subtest of **inferences** revealed no significant differences ($F = 1.42$, $df = 6/294$, $p = 0.206$) among the seven repeated measures as did the pretest/posttest analysis. Nevertheless, when these scores were graphed, as seen in Figure 2, a pattern was revealed which was different from the one indicated by the pretest/posttest scores. Since the ANOVA did not result in a significant difference between the means, no post hoc analysis was conducted.

The same method of analysis was employed for the subtest recognition of **assumptions**. This time the repeated measures ANOVA, like the pretest/posttest analysis, revealed a significant difference ($F = 2.35$, $df = 6/294$, $p = 0.039^*$). Nevertheless, even though the pretest/posttest scores would indicate a linear significant increase in this ability, the repeated measures design showed a decrease for the semester immediately following admission to the nursing major prior to the more or less linear increase (Figure 3).

Table 4.

Repeated Measures Analysis of Variance of WGCTA Subtest Scores

Source	<u>SS</u>	<u>df</u>	<u>MS</u>	<u>F</u>	<u>p</u>
Inference					
Between Groups	36.90	6	6.15	1.42	.206
Within Groups	1272.53	294	4.33		
Assumptions					
Between Groups	73.87	6	12.31	2.24	.039*
Within Groups	1612.42	294	5.48		
Deduction					
Between Groups	42.82	6	7.14	2.35	.031*
Within Groups	893.75	294	3.04		
Interpretation					
Between Groups	27.75	6	4.63	2.17	.046*
Within Groups	627.67	294	2.13		
Argument					
Between Groups	54.78	6	9.13	2.59	.019*
Within Groups	1037.79	294	3.53		

* $p < .05$

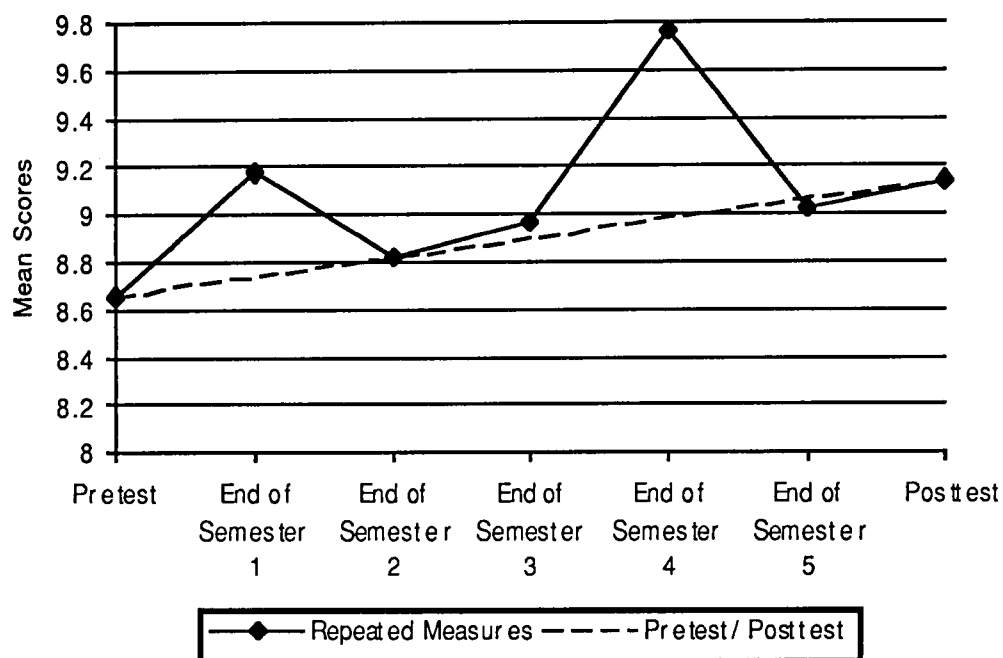


Figure 2. Repeated Measures Analysis of WGCTA Inference Scores

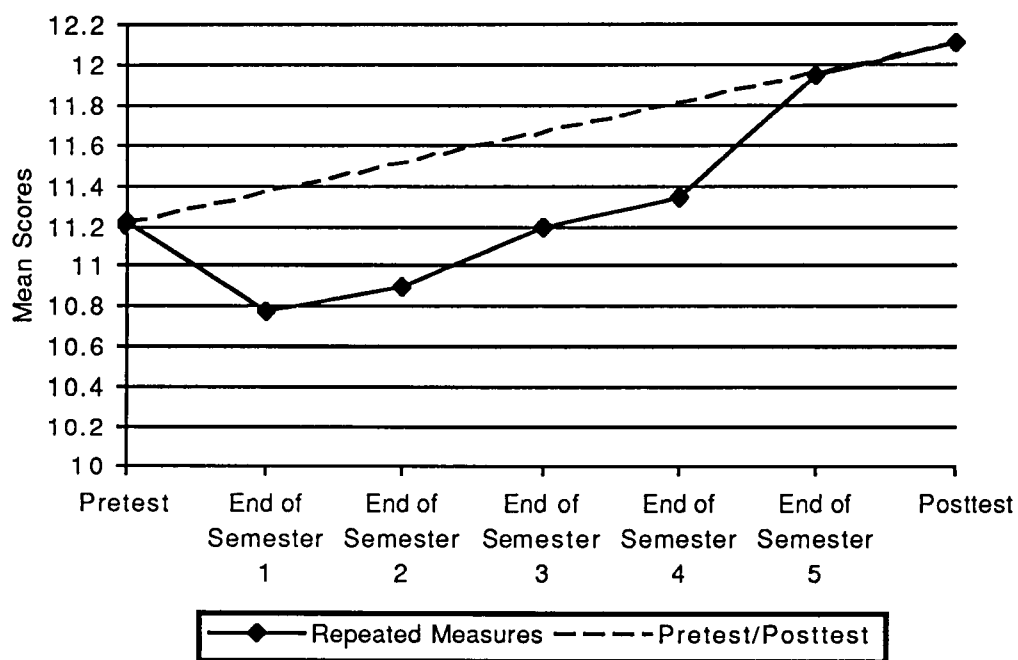


Figure 3. Repeated Measures Analysis of WGCTA Assumption Scores

However, Tukey's HSD post hoc analysis of this repeated measures ANOVA failed to find a significant pairwise difference between any of the means. This result might be explained by the fact that Tukey's HSD is a more conservative test than ANOVA. As such, it keeps the maximum joint alpha level constant regardless of the number of comparisons made and therefore places emphasis on limiting the probability of making a Type I error, that is stating that there was a difference when the difference was due merely to chance rather than a true difference. However, this analysis may have resulted in a Type II error by stating that no difference was found when, in fact, a true difference did exist.

Next the subtest for **deduction** was analyzed with a repeated measures ANOVA. This analysis also indicated a significant difference ($F = 2.35$, $df = 6/294$, $p = 0.031^*$) among the scores. Once more, when the repeated measures scores were graphed, a pattern different from the one for assumptions emerged. This pattern indicated that the scores decreased at the end of the first semester, the end of the third semester, and the end of the fifth semester. However, there were increases at the end of the second semester, the end of the fourth semester, and at the posttest. This pattern can be seen in Figure 4.

Once more a Tukey's HSD analysis of these findings was done which showed that the students' scores at the end of the fourth semester were significantly different from their scores at the end of the third semester ($q = 4.218^*$).

The repeated measures ANOVA for the subtest for **interpretation** indicated that unlike the pretest/posttest design, there were significant differences ($F = 2.17$, $df = 6/294$, $p = 0.046^*$) between the scores. When the repeated measures scores were graphed, a still different pattern of change emerged. This time the increases occurred at the end of the third semester and at the end of the fifth semester with decreases occurring

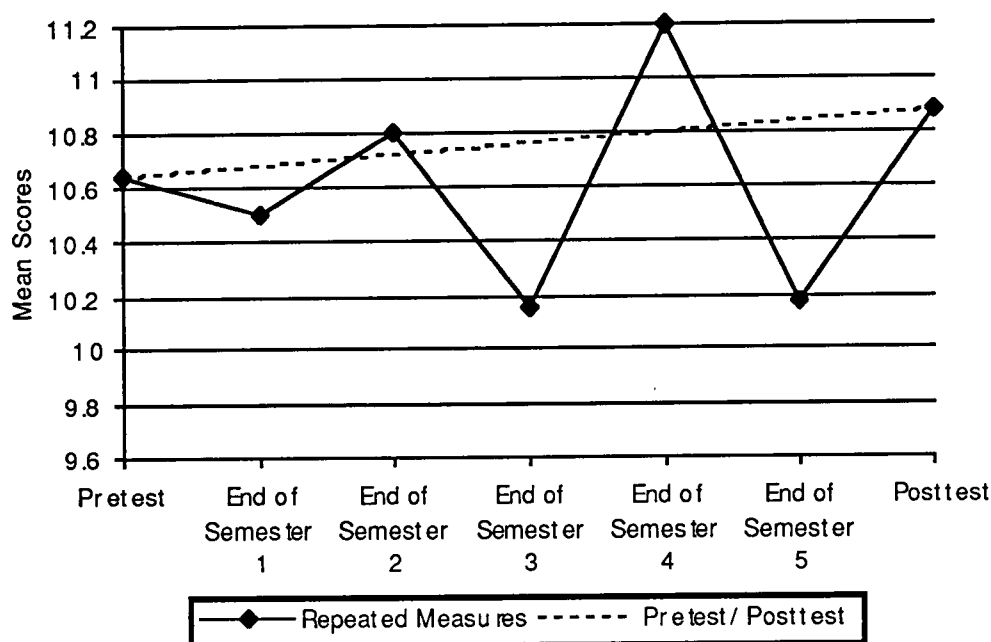


Figure 4. Repeated Measures Analysis of Deduction WGCTA Scores

at the end of the first semester, the end of the fourth semester, and at the posttest. Again this pattern was graphed and can be seen in Figure 5.

The post hoc analysis of this ANOVA using Tukey's HSD revealed that there were no pairwise comparisons which were significant. These results can be explained using the same logic used with the subtest for assumptions, in that, since Tukey's HSD is a more conservative test than ANOVA, it places more emphasis on limiting the probability of making a Type I error and as such did not show that any of the pairwise comparisons were significant. Once more, this analysis may have resulted in a Type II error.

The last subtest to be analyzed was that for **arguments**. For this subtest, the repeated measures ANOVA suggested that there were significant differences ($F = 2.59$ ($df = 6/294$, $p = 0.019^*$)) in the scores. Once more, when the repeated measures scores

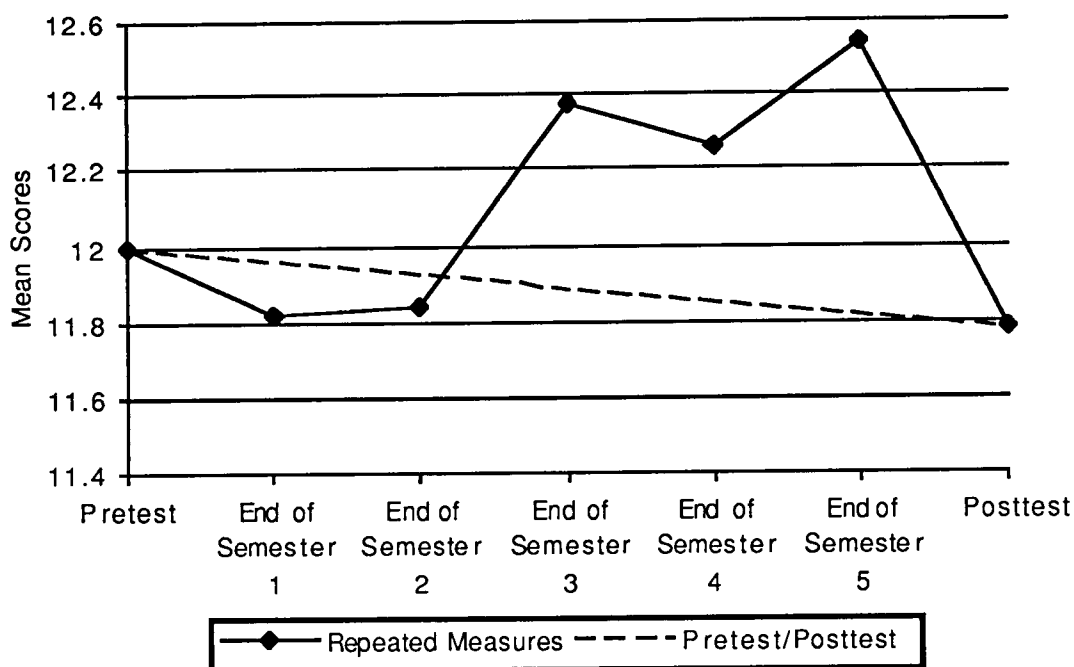


Figure 5. Repeated Measures Analysis of Interpretation WGCTA Scores

were graphed, a still different pattern of change appeared. This time the decreases occurred at the end of the first semester, the end of the third semester, the end of the fifth semester, and at the posttest. The increases were evident at the end of the second semester and at the end of the fourth semester. This pattern can be seen in Figure 6.

Again, Tukey's HSD post hoc analysis for these scores was conducted. The results indicated that the students' scores from the end of the second semester were significantly different from their scores at the end of the first semester ($q = 4.516^*$) as well as from their scores at the end of the third semester ($q = 4.215^*$).

In summary, the only significant difference from pretest to posttest for the subtest scores was the one for recognition of assumptions. The other pretest/posttest subtest scores showed non-significant changes with some increasing and others

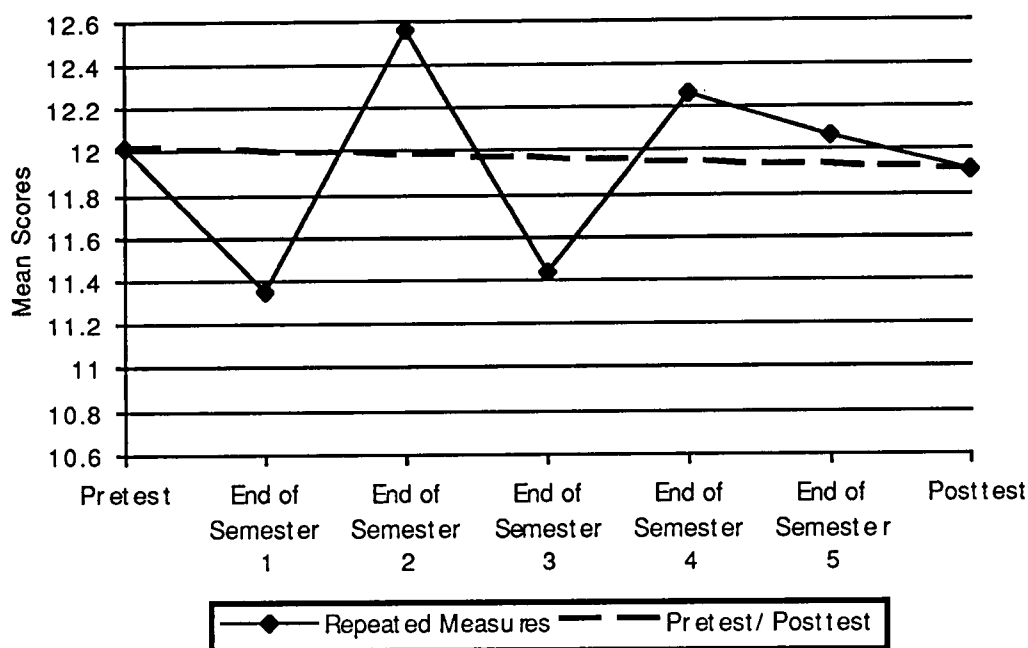


Figure 6. Repeated Measures Argument WGCTA Scores

decreasing. The repeated measures ANOVA for these scores revealed a significant difference between the means for all the subtest scores with the exception of inferences which showed no significant difference. The pattern of change in the scores for each of these subtests indicates a much different growth trajectory than did the pretest/posttest design. Thus, the answer to both the research questions is no, the pretest/posttest design does not lead to the same conclusions about changes in WGCTA total scores nor component scores as a repeated measures design.

CHAPTER V

DISCUSSION, CONCLUSION, AND RECOMMENDATIONS

Since critical thinking has become a discrete outcome measure for judging the quality of nursing programs, the issue of adequate and appropriate means of measurement of this ability is of paramount importance for nursing educators. The studies examining the effect of nursing curricula on critical thinking skills conducted thus far have yielded inconclusive and conflicting findings. What accounts for these findings? In order to discover potential answers to this question, the purpose of this study was to compare the effectiveness of a longitudinal repeated measures design model to a pretest/posttest design model in assessing changes in critical thinking abilities of a cohort of baccalaureate nursing students. This was done to determine if the research design was an important factor affecting the findings of such studies. The research questions guiding this study were: (1) Does a pretest/posttest design lead to the same conclusions about changes in total WGCTA critical thinking scores as a repeated measures design? and (2) Does a pretest/posttest design lead to the same conclusions about changes in component WGCTA critical thinking scores as a repeated measures design?

An existing data set consisting of Watson Glaser Critical Thinking Appraisal (WGCTA) scores obtained from 1992 through 1994 from a cohort of baccalaureate nursing students in a comprehensive university in the southeast was used to conduct this study. The WGCTA had been administered to students at the beginning of the first nursing course in the major and at the end of each subsequent semester of study. The first set of scores was used as pretest scores and the last set of scores was used as

posttest scores. The intervening scores were used together with the pretest and posttest scores as data for the repeated measures design.

With alpha set at 0.05, paired t-tests were used to analyze the data based on a pretest/posttest design, while a repeated measures analysis of variance (ANOVA) was used to analyze the data based on a longitudinal design. For ANOVAs which indicated a significant difference, a post hoc analysis using Tukey's Honestly Significant Difference (HSD) was computed to discover where the significant differences occurred. The independent variable for the study was the nursing curriculum while the dependent variable was the students' scores on the WGCTA instrument.

The answer to the first research question (does a pretest/posttest design lead to the same conclusions about changes in total WGCTA critical thinking scores as a repeated measures design) was no, the conclusions were not the same. The analysis of the data based on a simple pretest/posttest design indicated that there were no significant changes which occurred. These findings were supportive of some of the previous studies using this design (Sullivan, 1987; Bauwens & Gerhard, 1987; Kintgen-Andrews, 1988; Notarianni, 1991; Behrens, 1996), while, at the same time, refuted others (Berger, 1984; Gross, Takazawa, & Rose, 1987; Miller, 1992; Pepa, Brown, & Alverson, 1997). When the data were subjected to an analysis based on a repeated measures design however, there was evidence of significant changes throughout the course of study with scores increasing in some semesters and decreasing in others.

The answer to the second research question (does a pretest/posttest design lead to the same conclusions about changes in component WGCTA critical thinking scores as a repeated measures design) is also no, the conclusions are not the same. The analysis of the data based on a pretest/posttest design for the subtest scores suggested that no significant change had occurred with the exception of the subtest of recognition of assumptions which showed a significant increase. In contrast, use of the longitudinal

repeated measures design for the subtest scores revealed significant changes for four of the five the component scores including recognition of assumptions, deduction, interpretation, and evaluation of arguments. Only the fifth component score, inferences, showed no significant change.

When the two types of designs are compared, they lead to markedly different results. Using the pretest/posttest design would lead one to conclude that the nursing curriculum had had no significant effect on either the total WGCTA scores or the component scores with the exception of recognition of assumptions. On the other hand, using a repeated measures design demonstrates that the nursing curriculum did indeed have a significant effect on the total WGCTA scores as well as on the component scores, with the exception of inferences. When the results of these two designs are compared, the actual growth trajectory of critical thinking skills is one with both significant increases and decreases rather than being linear and flat.

Discussion

Comparison of this study with prior work might help to explain the inconclusive, mixed findings of previous studies attempting to measure the effect of nursing curricula on critical thinking abilities. Most previous studies used a pretest/posttest design. The pretest/posttest design of this study yields similar findings to other pretest/posttest studies, i.e. that nursing curricula had had no significant effect on critical thinking. However, when the repeated measures design is compared to pretest/posttest designs, the opposite notion is supported, i.e., that nursing curricula have a significant effect on critical thinking abilities and this effect appears to be non-linear and quite dynamic. In addition the results from this type of design seems to refute the assumption that nursing curricula have a linear cause and effect on critical thinking abilities of students. A possible explanation for this finding is that there may well have been periods in previous pretest/posttest studies when the nursing curricula had had a

significant positive as well as a negative effect on critical thinking but, because of the research design method used, these effects were obscured leading to conclusions which were either invalid or inconclusive.

Previous studies utilizing a pretest/posttest design have resulted in a limited view of the development of critical thinking abilities. In addition to the inherent flaws in that type of design, it has also prevented researchers from determining what happened to this ability during the time the students were in the “black box” of time called the course of study. However, when this “black box” is exposed through use of a more appropriate design, faculty are provided with a rich and full data base with which to assess this outcome. It then follows that faculty would have valid information with which to develop, maintain, or revise curricula as they deem necessary in order for the goal of increased critical thinking to be achieved.

A growth trajectory of critical thinking skills provides a broader and more extensive picture for educators in understanding students’ progress. The repeated measures design, rather than a pretest/posttest design, is better suited to providing this picture. If nursing faculty are truly interested in determining the effect that curricula have on the development of critical thinking abilities of students, then a repeated measures design should be used. This type of design allows for collection of data at multiple points and thus yields enough data about the phenomenon for valid conclusions to be drawn and appropriate decisions to be made. Additionally, the repeated measures design provides a mechanism for researchers to answer the question of not only how students fared from one point to another in their critical thinking skills, but how they got from one point to another, a finding which is totally obscured in the pretest/posttest design.

Because the authors of the instrument have stated that the subtests should not be used to determine individual attainment of these skills due to the small number of items

in each subtest, previous studies have not specifically looked at students' scores in this area. However, the authors go on to state that the subtests scores could be used to analyze these abilities for a class or larger group. Consequently, it would seem that if a repeated measures design was used so that several evaluations of these abilities were assessed, then these subtest scores could be used to show where in the curriculum certain of the skills were enhanced while others were negatively affected. This information could then be used to determine where changes in the curriculum related to component critical thinking abilities needed to be made that would result in positive outcomes relative to critical thinking.

The repeated measures design revealed a pattern of both increases and decreases in critical thinking abilities. Several factors may have been operating. The first centers around faculty. The ability of faculty to teach critical thinking skills may vary. Teaching methods used by the faculty in the various semesters may make a difference in the critical thinking skills achieved by students, with some supporting the attainment of these skills and others hindering them. Additionally, the teaching experience of faculty may play an important role. Neophyte faculty may be hesitant to use teaching methods which could be viewed by their colleagues as unconventional, and thus may rely on the traditional lecture which may not be particularly successful in teaching students to think (DeYoung, 1990).

The specific course content could make a difference in whether or not students increased their critical thinking abilities. Some courses lend themselves more to the teaching and exercising of critical thinking skills than others. For instance, some nursing courses are aimed primarily at teaching students psychomotor skills (e.g., physical assessment, nursing fundamentals) where there is little need for students to recognize assumptions or evaluate arguments. Others, such as nursing issues or

nursing ethics and especially clinical courses, require that students explore and reflect upon the content while using all of the skills identified by Watson and Glaser.

Both the total scores and all of the subtest scores, with the exception of inference, decreased after the first semester of study in the nursing major. The first nursing course was a traditional "fundamentals" course with emphasis on psychomotor skills rather than decision making and problem solving. Students learn procedures in more of a rote fashion and are expected to perform these procedures in a specific way under the supervision of faculty. This narrow focus encourages students who are neophytes in the profession and who are afraid of making errors to seek concrete, right/wrong, and black/white answers.

Conversely, during the fourth semester when students have a preceptor experience in which they are supported and encouraged to assume the role and responsibilities of a practicing nurse, both the total scores and the subtest scores increase with the exception of interpretation. Then, in the following semester when they are again in the role of student, total scores as well as the subtest scores of deduction, argument, and inference decline. It can be postulated that the independent practice of nursing mandates critical thinking so that when students are assuming a role more of a practicing nurse than of a student, they exercise these abilities. When again placed in a role of student these abilities are suppressed.

Many factors relating specifically to the students might account for the increases and decreases revealed by the repeated measures design. The number and type of non-nursing as well as nursing courses in which the students are enrolled each semester varies and could be having an effect, with some supporting and others discouraging critical thinking. The effect of fatigue on their level of performance is likewise undetermined. Fatigue could be due to both the academic load carried by the student as well as the fact that all of the Watson Glaser instruments were administered during the

last week of classes or during the week of final examinations with the exception of the first one. At this time during the semester students are frequently becoming tired from the rigors of course work and may not have given their best effort at each measurement session. It can be assumed that as fatigue increases thinking abilities decrease.

Motivation may have also affected the students' performance. Since faculty did not express an expectation for level of performance or consequence for poor performance, the students may or may not have given their best effort each time the test was offered.

Conclusion

Based on the findings of this study, the conclusion which can be reasonably drawn is that a repeated measures design is much better than a pretest/posttest design for assessing changes in critical thinking. Nevertheless, care must be taken in extrapolating these results beyond the current study due to the small sample size which limits the external validity.

Implications for Practice

Educators, educational organizations, and accrediting bodies continue to assert that critical thinking is an essential attribute for college graduates. As a result, faculties and administrators must expend the resources necessary to determine if their curricula, in fact, do lead to the acquisition of this ability. Part of this expenditure must include the selection of an appropriate assessment design which leads to valid findings rather than a method which is most convenient. The findings of this study seem to indicate that a repeated measures design is the most comprehensive and powerful method with which to study this ability. Therefore, the implication for practice is that a longitudinal repeated measures design should be used to assess critical thinking abilities rather than a pretest/posttest design. The repeated measures design requires a commitment of time and money, both of which are at a premium in an educational setting. However, if

adequate amounts of resources are not spent in this regard, programs will not have an accurate view of the effect of their curricula on critical thinking skills.

With results from a repeated measures design, faculty can identify which parts, if any, of the curriculum are more successful in fostering critical thinking and which aspects inhibit this skill. They could then determine if the curriculum needs to be restructured or reorganized to facilitate critical thinking. Further, course assignments and activities as well as evaluation methods could be examined and integrated in light of the overall goal of increased critical thinking.

Since it was determined that various factors related to students may have had an impact on the findings of the repeated measures design, several implications for practice are indicated. Faculty should encourage students to become partners in their own learning and to seek opportunities and activities which enhance their critical thinking abilities as well as to value opportunities to assess their performance in relation to this outcome. Timing of these evaluation measures also needs to be considered. Times for test administration need to be selected when students are less likely to be fatigued.

Recommendations for Further Research

Although the findings of this study do not conclusively determine that a repeated measures design is better than a pretest/posttest design, they do illustrate the need for further study on the impact of research design on study results. One of the first steps to accomplish this task would be to replicate the current study using a larger sample from many institutions with random selection of participants in order to increase external validity. This study should also be repeated with students who are pursuing majors other than nursing to determine if similar results would be obtained. The use of a comparison group would assist one to rule out rival explanations such as that the findings were limited to nursing students. Additionally, the study should be replicated

using a different instrument to measure critical thinking skills in order to determine if the results were merely an artifact of the instrument used.

REFERENCES

References

Aaronson, L. S., & Kingry, M. J. (1988). A mixed method approach for using cross-sectional data for longitudinal inferences. Nursing Research, 37(3), 187-189.

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

American Association of Colleges of Nursing. (1986). Essentials of college and university education for professional nursing. Washington, DC: Author.

American Nurses' Association Committee on Education. (1965). American Nurses' Association first position on education for nursing. American Journal of Nursing, 65(12), 104-111.

Association of American Colleges. (1985). Integrity in the college curriculum: A report to the academic community. Washington, DC: Author.

Astin, A. W. (1993). Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education. Phoenix, AZ: Oryx Press.

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), Longitudinal research in the study of behavior and development (pp. 1-40). New York: Academic Press.

Bandman, E. L., & Bandman, B. (1988). Critical thinking in nursing. Norwalk, CT: Appleton & Lange.

Baron, J. B., & Sternberg, R. J. (1987). Teaching thinking skills: Theory and practice. New York: W. H. Freeman Publishing.

Bauwens, E. E., & Gerhard, G. G. (1987). The use of the Watson-Glaser Critical Thinking Appraisal to predict success in a baccalaureate nursing program. Journal of Nursing Education, 26(7), 278-281.

Behrens, P. J. (1996). The Watson-Glaser Critical Thinking Appraisal and academic performance of diploma school students. Journal of Nursing Education, 35(1), 34-36.

Berger, A. (1985). Review of the Watson-Glaser Critical Thinking Appraisal. In J. V. Mitchell, Jr., (Ed.), The ninth mental measurements yearbook (Vol 2, pp. 1692-1693), Lincoln, NE: University of Nebraska Press.

Berger, M. (1984). Clinical thinking ability and nursing students. Journal of Nursing Education, 23(7), 306-308.

Bevis, E. O., & Watson, J. (1989). Toward a caring curriculum: A new pedagogy for nursing. New York: National League for Nursing.

Boyer, E. L. (1987). College: The undergraduate experience in America. New York: Harper & Row, Publishers.

Brigham, C. F. (1989). Critical thinking skills in nursing students progressing through a nursing curriculum [CD-ROM]. Abstract from: ProQuest File: Dissertation Abstracts Item: 8918496.

Brooks, K. L., & Shepherd, J. M. (1990). The relationship between clinical decision-making skills in nursing and general critical thinking abilities of senior nursing students in four types of nursing programs. Journal of Nursing Education, 29(9), 391-399.

Brooks, K. L., & Shepherd, J. M. (1992). Professionalism versus general critical thinking abilities of senior nursing students in four types of nursing curricula. Journal of Professional Nursing, 8(2), 87-95.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Company.

Case, B. (1994). Walking around the elephant: A critical-thinking strategy for decision making. The Journal of Continuing Education in Nursing, 25(3), 101-109.

College Board. (1983). Academic preparation for college. New York: College Entrance Examination Board.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design & analysis for field settings. Chicago: Rand McNally College Publishing Company.

Currier-Dagrosa, C. W. (1986). The effects of summer work study programs upon the problem solving skills of baccalaureate nursing students [CD-ROM]. Abstract from: ProQuest File: Dissertation Abstracts Item: 8619106.

Danielson, D. J. (1985). Variables which influence professional socialization in the nursing student [CD-ROM]. Abstract from: ProQuest File: Dissertation Abstracts Item: 8524542.

DeYoung, S. (1990). Teaching nursing. New York: Addison-Wesley Nursing.

Dressel, P. L., & Mayhew, L. B. (1954). General education: Explorations in evaluation. Washington, DC: American Council on Education.

Ennis, R. H. (1962). A concept of critical thinking. Harvard Educational Review, 32(1), 81-111.

Ennis, R. H. (1985). A conception of critical thinking with some curriculum suggestions. In J. Hoaglund (Ed.), Conference 85 on critical thinking (pp. 13-34). Newport News, VA: Christopher Newport College Press.

Ewell, P. T. (1988). Outcomes, assessment, and academic improvement: In search of usable knowledge. In J. C. Smart (Ed.), Higher education: Handbook of theory and research: Vol. 4, (pp. 53-104). New York: Agathon Press, Inc.

Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations. Newark, DE: American Philosophical Association. (ERIC Document Reproduction Services No. ED 315 423).

Facione, N. C., Facione, P. A., & Sanchez, C. A. (1994). Critical thinking disposition as a measure of competent clinical judgment: The development of the California Critical Thinking Disposition Inventory. Journal of Nursing Education, 33(8), 345-350.

Ferguson, G. A., & Takane, Y. (1989). Statistical analysis in psychology and education (6th ed.). New York: McGraw-Hill Book Company.

Fleeger, R. L. (1987). Critical thinking and moral reasoning behavior of baccalaureate nursing students [Abstract]. Proceedings of the Fifth Annual Research in Nursing Education Conference, 5, 34.

Fraenkel, J. R., & Wallen, N. E. (1996). How to design and evaluate research in education (3rd ed.). New York: McGraw-Hill, Inc.

Franchak, S. J., Franken, M. E., Subisak, J., Kapes J. T., & Sheldon, M. T. (1980). Specifications for longitudinal studies. Columbus, OH: The National Center for Research in Vocational Education.

Frederickson, K., & Mayer, G. G. (1977). Problem solving skills: What effect does education have?. American Journal of Nursing, 77(7), 1167-1169.

Freed, M. N., Ryan, J. M., & Hess, R. K. (1991). Handbook of statistical procedures and their computer applications to education and the behavioral sciences. New York: Macmillan Publishing Company.

Gross, Y. T., Takazawa, E. S., & Rose, C. L. (1987). Critical thinking and nursing education. Journal of Nursing Education, 26(8), 317-323.

Grossman, S. C. (1985). Factors associated with the clinical competence of students enrolled in generic baccalaureate nursing programs: Implications for program management and design [CD-ROM]. Abstract from: ProQuest File: Dissertation Abstracts Item: 8516188.

Gunning, C. S. (1981). Relationships among field independence, critical thinking ability, and clinical problem-solving ability of baccalaureate nursing students [CD-ROM]. Abstract from: ProQuest File: Dissertation Abstracts Item: 8128633.

Harmer, B. (1928). Textbook of the principles and practice of nursing (2nd ed.). New York: Macmillan Company.

Heaslip, P. (1994). Critical thinking in nursing curricula: Curriculum revolution in action. Educational Vision, 2(2), 32-33.

Helmstadter, G. C. (1985). Review of the Watson-Glaser Critical Thinking Appraisal. In J. V. Mitchell, Jr., (Ed.), The ninth mental measurements yearbook (Vol 2, pp. 1693-1694), Lincoln, NE: University of Nebraska Press.

Jones, J. T. (1985). A comparison of critical thinking ability of graduating baccalaureate and associate degree nursing students [CD-ROM]. Abstract from: ProQuest File: Dissertation Abstracts Item: 8423490.

Jones, S. A., & Brown, L. N. (1991). Critical thinking: Impact on nursing education. Journal of Advanced Nursing, 16, 529-533.

Kataoka-Yahiro, M., & Saylor, C. (1994). A critical thinking model for nursing judgment. Journal of Nursing Education, 33(8), 351-356.

Kerlinger, F. N. (1973). Foundations of behavioral research (2nd ed.). New York: Holt, Rinehart and Winston, Inc.

Keselman, H. J., & Keselman, J. C. (1993). Analysis of repeated measurements. In L. K. Edwards (Ed.), Applied analysis of variance in behavioral science (pp. 105-145). New York: Marcel Dekker, Inc.

Ketefian, S. (1981). Critical thinking, educational preparation, and development of moral judgment among selected groups of practicing nurses. Nursing Research, 30(2), 98-103.

Kintgen-Andrews, J. (1988). Development of critical thinking: Career ladder P.N. and A.D. nursing students, prehealth science freshman, generic baccalaureate sophomore nursing students. Resources in education, 24(1). (ERIC Document Reproduction Service No. ED 297 153).

Klaassens, E. L. (1988). Improving teaching for thinking. Nurse Educator, 13(6), 15-19.

Kokinda, M. A. (1989). The measurement of critical thinking skills in a selected baccalaureate nursing program [CD-ROM]. Abstract from: ProQuest File: Dissertation Abstracts Item: 9004798.

Kramer, M. K. (1993). Concept clarification and critical thinking: Integrated processes. Journal of Nursing Education, 32(9), 406-414.

Leedy, P. D. (1985). Practical research: Planning and design (3rd ed.). New York: Macmillan Publishing Company.

Leedy, P. D. (1989). Practical research: Planning and design (4th ed.). New York: Macmillan Publishing Company.

Levin, J. R., Marascuilo, L. A., & Hubert, L. J. (1978). Nonparametric randomization tests. In T. R. Kratochwill (Ed.), Single subject research: Strategies for evaluating change (pp. 167-196). New York: Academic Press.

Light, R. J., Singer, J. D., & Willett, J. B. (1990). By design: Planning research on higher education. Cambridge, MA: Harvard University Press.

Lindeman, C. (1995). Emerging environment for nursing education and practice examined during two-year Vision Campaign. New York: National League for Nursing.

Lynch, M. H. (1988). Critical thinking: A comparative study of baccalaureate and associate degree nursing students (Doctoral dissertation, Peabody College for

Teachers of Vanderbilt University, 1988). Dissertation Abstracts International, 49, 2157.

Matthews, C., & Gaul, A. (1979). Nursing diagnosis from the perspective of concept attainment and critical thinking. Advances in Nursing Science, 2(1), 17-26.

Maynard, C. A. (1996). Relationship of critical thinking ability to professional nursing competence. Journal of Nursing Education, 35(1), 12-18.

McManus, L. (1949). The effect of experience on nursing achievement. New York: Columbia University, Teachers College, Bureau of Publications.

McMillian, J. H. (1987). Enhancing college students' critical thinking: A review of studies. Research in Higher Education, 26(1), 3-29.

McPeck, J. E. (1981). Critical thinking and education. New York: St. Martin's Press.

McPeck, J. E. (1990). Teaching critical thinking: Dialogue and dialectic. New York: Routledge, Chapman and Hall, Inc.

Miller, M. A. (1992). Outcomes evaluation: Measuring critical thinking. Journal of Advanced Nursing, 17, 1401-1407.

Miller, M. A., & Malcolm, N. S. (1990). Critical thinking in the nursing curriculum. Nursing & Health Care, 11(2), 66-73.

Mitchell, M., & Jolley, J. (1988). Research design explained. New York: Holt, Rinehart & Winston, Inc.

National Education Goals Panel. (1991). Executive summary: The National Education Goals Report, 1991: Building a nation of learners. Washington, DC: Author.

National Institute of Education. (1984). Involvement in learning: Realizing the potential of American higher education. Washington, DC: U. S. Government Printing Office.

National League for Nursing. (1983). Criteria for baccalaureate and higher degree programs in nursing. New York: Author.

National League for Nursing. (1987). Characteristics of baccalaureate education in nursing. New York: Author.

National League for Nursing. (1989). Criteria for the evaluation of baccalaureate and higher degree programs in nursing (6th ed.). New York: Author.

National League for Nursing. (1992). Criteria and guidelines for the evaluation of baccalaureate and higher degree programs in nursing (7th ed.). New York: Author.

Notarianni, M. A. (1991). An investigation of the critical thinking ability of associate and baccalaureate degree nursing students (Doctoral Dissertation, Widener University, 1991). Dissertation Abstracts International, 52, 2503.

Pardue, S. F. (1987). Decision-making skills and critical thinking ability among associate degree, diploma, baccalaureate, and master's-prepared nurses. Journal of Nursing Education, 26(9), 354-361.

Pascarella, E. T., & Terenzini, P. T. (1991). How college affects students. San Francisco: Jossey-Bass Publishers.

Paul, R. (1984). Critical thinking: Fundamental for a free society. Educational Leadership, 41, 4-14.

Paul, R. W. (1993) Critical thinking: What every person needs to survive in a rapidly changing world (Rev. 3rd ed.). Santa Rosa, CA: Foundation for Critical Thinking.

Pepa, C. A., Brown, J. M., & Alverson, E. M. (1997). A comparison of critical thinking abilities between accelerated and traditional baccalaureate nursing students. Journal of Nursing Education, 36(1), 46-48.

Poole, M. H. (1989). An assessment of critical thinking skills in nursing students (Doctoral Dissertation, University of Houston-University Park, 1989). Dissertation Abstracts International, 51,429.

Rane-Szostak, D., & Robertson, J. F. (1996). Issues in measuring critical thinking: Meeting the challenge. Journal of Nursing Education, 35(1), 5-11.

Rockefeller Commission on the Humanities. (1980). The humanities in American life. Berkeley, CA: University of California Press.

Scolaveno, M. A. (1981). Problem-solving ability of senior nursing students in three program types [CD-ROM]. Abstract from: ProQuest File: Dissertation Abstracts Item: 8120858.

Schank, M. J. (1990). Wanted: Nurses with critical thinking skills. The Journal of Continuing Education in Nursing, 21(2), 86-89.

Sheskin, D. J. (1997). Handbook of parametric and nonparametric statistical procedures. Boca Raton, FL: CRC Press.

Siegel, H. (1988). Educating reason: Rationality, critical thinking, and education. New York: Routledge.

Skinner, S. B. (1976). Cognitive development: A prerequisite for critical thinking. The Clearing House-A journal for modern junior and senior high schools, 49(7), 292-299.

Soefje, L. T. (1985). Differences in selected aspects of professionalism between registered nurse students entering and those exiting baccalaureate degree programs in nursing [CD-ROM]. Abstract from: ProQuest File: Dissertation Abstracts Item: 8514293.

Stewart, I. M. (1944). The education of nurses. New York: Macmillan Company.

Sullivan, E. J. (1987). Critical thinking, creativity, clinical performance, and achievement in RN students. Nurse Educator, 12(2), 12-16.

Tabachnick, B. G., & Fidell, L. S. (1983). Using multivariate statistics. New York: Harper & Row, Publishers.

Task Force on Education for Economic Growth of the Education Commission of the States. (1983). Action for excellence. Denver, CO: Education Commission of the States.

Tiessen, J. B. (1987). Critical thinking and selected correlates among baccalaureate nursing students. Journal of Professional Nursing, 3(2), 118-123.

Waite, R. M. (1989). A measurement of critical thinking in senior baccalaureate nursing students [CD-ROM]. Abstract from: ProQuest File: Dissertation Abstracts Item: 9009974.

Watson, G., & Glaser, E. M. (1964). Critical thinking appraisal manual. New York: Harcourt Brace & World.

Watson, G., & Glaser, E. M. (1980). Watson-Glaser critical thinking appraisal manual. San Antonio, TX: The Psychological Corporation, Harcourt Brace Jovanovich, Inc.

White, N. E., Beardslee, N. Q., Peters, D., & Supples, J. M. (1990). Promoting critical thinking skills. Nurse Educator, 15(5), 16-19.

Yura, H. (1978). Who is the product of the baccalaureate program? Curriculum process for developing or revising a baccalaureate nursing program. New York: National League for Nursing.

APPENDICES

APPENDIX A: PERMISSION TO USE DATA SET

THE UNIVERSITY OF TENNESSEE
AT CHATTANOOGA



School of Nursing
615 McCallie Avenue
Chattanooga, TN 37403
(423) 755-4644
FAX: (423) 755-4668

MEMORANDUM

TO: Barbara Norwood, RN, MSN

FROM: Cherry Guinn, RN, EdD, Project Director *CG*

RE: Permission to use data set

DATE: September 6, 1996

You have my permission to use the data set from my research project, "Critical Thinking in the UTC Nursing Curriculum". Best wishes for a successful adventure with your dissertation.

APPENDIX B: IRB APPROVAL

#20

FORM A

IRB # _____

Certification of Exemption from Review by Full Committee
for Research Involving Human Subjects

Date received in OR _____

- A. PRINCIPLE INVESTIGATOR(s) and/or CO-PI(s): (For student projects, list both the student and the advisor.)
Barbara R. Norwood
Dr. Norma T. Mertz, Advisor
- B. DEPARTMENT/UNIT: Leadership Studies, College of Education
- C. COMPLETE MAILING ADDRESS AND PHONE NUMBER OF PI(s)
and CO-PI(s):
Barbara R. Norwood, 418 Peyton Dr., Hixson, TN., 37343 423-843-0795
Dr. T. Mertz, 5614 Far Vista Ln., Knoxville, TN., 37914 423-974-6150
- D. TITLE OF PROJECT: Critical Thinking in Nursing Students: A Comparison of Designs
- E. EXTERNAL FUNDING AGENCY AND ID NUMBER (if applicable): NA
- F. GRANT SUBMISSION DEADLINE (if applicable): NA
- G. STARTING DATE: 8/97 (NO RESEARCH MAY BE INITIATED UNTIL
CERTIFICATION IS GRANTED)
- H. ESTIMATED COMPLETION DATE (Include all aspects of research and final write-up): 12/97
- I. Objective(s) of Project (Use additional page, if needed.): To compare the effectiveness of a longitudinal repeated measures design model to a pretest/posttest design model in
- II. Subjects (Use additional page, if needed.): Data set consisting of critical thinking scores from a cohort of students (n=63) who progressed through a baccalaureate
- III. Methods or Procedures (Use additional page, if needed.): The study is comparative in design, allowing one to compare the results of using different methodologies
- IV. CATEGORY(ies) FOR EXEMPT RESEARCH PER 45 CFR 46 (see reverse side for categories): 4

CERTIFICATION: The research described herein is in compliance with 45 CFR 46.101(b) and presents subjects with no more than minimal risk as defined by regulations.

Principal Investigator BARBARA R. NORWOOD Barbara R. Norwood 7-17-97
Name Signature Date

Student Advisor Norma T. Mertz Norma T. Mertz 7/17/97
Name Signature Date

Dept. Review Comm. Chair Jeffery P. Aker JEFFERY P. AKER 7/17/97
Name Signature Date

APPROVED: RADY BOGUEZ RADY BOGUEZ 7/18/97
Dept. Head Name Signature Date

Earlly Revins - Carol Kasoon 8/27/97

I N T E R

College of Education

O F F I C E

MEMO

To: ✓ Ms. Barbara Norwood
 418 Peyton Dr.
 Hixson, TN 37343

Dr. Norma Mertz
CA227
Campus

From: Dr. Carol Kasworm

Date: August 29, 1997

Subject: Approval of Form A Review - Use of Human Subjects in Research

I authorize approval of your Form A for Compliance with Human Subject Research Guidelines for the Project: "Critical Thinking in Nursing Students: A Comparison of Designs." Thank you for your additional information to complete this review.

I wish you well on your research efforts. Attached is the signed permission form. If you have any questions, please do feel free to contact me at (423)974-2272 or ckasworm@utk.edu.

Good Luck!

c: Ms. Brenda Lawson
 Office of Compliance

VITA

Barbara Rohrer Norwood was born in Knoxville, Tennessee, where she attended public schools. In 1967 she entered the University of Tennessee at Knoxville where she completed pre-nursing requirements prior to her admission to the University of Tennessee Medical Units, Memphis, Tennessee. She was awarded a Bachelor of Science in Nursing Degree in 1971. In December 1978, she received a Masters of Science in Nursing, as a Clinical Specialist in Adult Health with a functional role in Education from the University of Tennessee, Knoxville.

Her professional background has included working as a staff nurse and a head nurse at Baroness Erlanger Hospital, Chattanooga, Tennessee. The majority of her career has been spent as a nursing educator in the baccalaureate nursing program in the School of Nursing, at the University of Tennessee, Chattanooga. She has maintained an active clinical practice in many of the hospitals in the Chattanooga area.

She entered the doctoral program in the College of Education in Leadership Studies, at the University of Tennessee, Knoxville, and received her Doctor of Education degree in May, 1998.