CHAPTER 5
DISCUSSION

Introduction

The purpose of this study was to determine if computer-assisted self-regulated learning strategies are effective tools to improve nursing student performance in science courses. Any improved performance may lead to decreased anxiety and increased self-efficacy, which can then lead to improved performance. The study accommodated all three phases of self-regulation from Schunk and Zimmerman (1998). The study design used survey techniques in the performance phase and a computer-assisted environment to provide feedback during the modeling/performance phase. Finally, the reflective phase of self-regulation was implemented through use of a comparison of exam scores to a final exam score. This chapter addresses the research questions of this study.

Study Sample

The study group sample population consisted of 85 graduate nursing students attending a mid-western university and enrolled in a graduate pathophysiology course. Seventy-five of the nursing students (88.2%) completed the research survey and agreed to allow their exam scores to be used in this research study. All of the students were taking the course for the first time.

Of the participating students, 12 held master’s degrees and were returning to school for a post-master’s certificate or to complete a degree to transition from a clinical nurse specialist to a nurse practitioner. The remaining 63 students had completed a bachelor’s degree in nursing, a requirement prior to entering the master’s program.
Research Questions

Can instruction in computer-assisted self-regulated learning techniques be used to improve student outcomes in a science course?

1. What are the time budget demands on a graduate nursing student? Do time budget demands have an effect on exam grades in a science course?

2. What importance to practice do nursing graduate students place on science course objectives both knowledge level (undergraduate) and application level (graduate)?

3. How does providing computer-assisted tutorials change student performance on exams as compared to pencil/paper worksheets?

4. Do elaborate feedback answers affect student performance on exams as compared to verification (correct/incorrect) type feedback?

5. Do students retain concepts taught through use of computer tutorials differently than concepts taught using paper and pencil worksheets as measured through a comprehensive final exam?

Research Question One

What are the time budget demands on a graduate nursing student? Do time budget demands have an effect on exam grades in a science course?

Adult learners have identified major time constraints that results from balancing home and family demands, work demands, and school demands (Kirby et al., 2004). The time constraints or time-budget demands on the study participants were categorized into three distinct groups: home constraints, work constraints and school constraints. Students were surveyed at the beginning of the course to determine the demands on their time related to these three constraints.
Most of the students in this study had multiple demands on their time from these three constraints. Since significant differences were found between the computer exam scores and the paper and pencil exam scores, an analysis of covariance (ANCOVA) was conducted on each time constraint to determine the potential effects of student time constraints on these scores. The paper and pencil exam scores were entered as the covariate, the computer exam scores were the dependent variable, and the time data were used as the fixed factor for each analysis.

**Home Constraints**

The student questions on the survey for home constraints asked students to relate how many children they had living at home. The analysis was two-fold, one analysis considering if the student had children or not and a second analysis considering how many children the student had living at home.

The number of students without children at home was 33 (44%) and the number with children at home was 42 (56%). The calculated mean score for students with children at home was 87.12 and for those without children at home was 87.36. The adjusted mean score for students with children at home was 86.97 and for those without children at home was 87.55. The analysis of covariance did not reveal any significant differences in exam scores when controlling for the covariate having children, F (1, 72) = 0.204, p = 0.653, partial $\eta^2 = 0.003$.

The second analysis considered number of children. Ten students (13.3%) had one child at home. Twenty students (26.7%) had two children at home. Eleven students (14.7%) had three children at home and one student had four children at home. The estimated marginal means for this grouping ranged from a low of 85.74 for those with two children to a high of 88.46 for the group with three or more children. The ANCOVA indicated no significant difference in mean exam scores from number of children, F (3, 70) = 0.733, p = 0.536, partial $\eta^2 = 0.030$. 
Two students indicated anecdotally that they were caring for two elderly parents at home. The time constraints relating to caring for elderly parents for these students were not considered, however.

**Work Constraints**

Although the course covers all body systems, some students may have different exam scores based upon their clinical experiences, years away from specific content areas, and time in their present jobs. Nearly all members of the study group worked (96%). Three types of work questions were asked in the survey: working and non-working, work setting or environment, and hours working per week. Analysis of working compared to non-working was not completed since only three members (4%) of the research participants did not work.

The first work-related analysis was to assess different work environments for the students. Student responses indicated four basic types of work experiences including those working: in acute care, including those who teach nursing in an acute care clinical environment (60%); working in an out-patient clinic or working as a school nurse (24%); working in psychiatric or mental health settings (9%); and those not working.

The adjusted marginal means ranged from 86.78 for acute care nurses to 90.61 for the nurses not working. The outpatient clinic group and the psychiatric mental-health groups fell between this range with adjusted means of 87.04 and 88.73 respectively. The ANCOVA for work experience indicated no significant difference, $F (3, 70) = 0.703$, $p = 0.533$, partial $\eta^2 = 0.029$.

All but three students (96%) work at least part time. Five students (6.7%) work less than 20 hours per week or less than half time. Four students (5.3%) worked from 20 – 30 hours per week. The majority of students work more than 30 hours per week. The majority of students
were working full time or more. The ANCOVA of hours spent working and the potential effects on test scores indicated no significant difference between the scores on the exams based upon hours working per week, $F (3, 70) = 0.477, p = 0.699$, partial $\eta^2 = 0.020$.

The final work constraint was to assess the independent and dependent variables as a function of the number of years in their position at work. Seven students (9.3%) were starting new positions at the beginning of the course. Sixteen students (20.3%) had been in their position one year or less. Fifteen students (20%) had been in their position two years or less and fourteen (18.7%) were three years in their position. The other students indicated from four to seventeen years in their current positions. The adjusted means for the number of years in the current job ranged from 83.98 for 3 years or less up to 89.68 for zero years of experience. The ANCOVA yielded an $F (4, 69)$ of 1.803 ($p = 0.138$, partial $\eta^2 = 0.095$); therefore, no significant difference between the variables was detected as a result of number of years in their position at work.

**School Constraints**

Questions about school constraints asked in the survey and considered in the analysis included: previous degrees held, number of classes being taken, distance traveled to attend class, and potential assistance from a study partner. The first consideration for school constraints was the degree the student held at the time of the course. All students in the course were required to have a bachelor of science in nursing prior to admittance to the graduate nursing program. However, 12 students (16%) held master’s degrees in nursing prior to attending the course. The homogeneity-of-slopes assumption was violated for the dependent variable, so a two-way repeated measures ANOVA was completed for this analysis. The results of the ANOVA were not significant, $F (1, 73) = 0.659$, $p = 0.419$, partial $\eta^2 = 0.009$. 
Fourteen students (18.7%) were taking only the pathophysiology course. The other students ranged from taking one to four additional courses. Although the credit hours were requested, many of the students failed to identify how many credit hours were being taken so this was not considered in the analysis. Some of the nursing graduate elective courses at this university are one credit hour and can range up to three credit hours. The adjusted means for this analysis ranged from a low of 85.86 for those students taking two or more additional courses to 88.12 for the students taking one additional course. The ANCOVA for courses enrolled indicated no significant effects on the exam score difference, $F(2, 71) = 1.241, p = 0.295$, partial $\eta^2 = 0.034$.

Students often have to attend classes multiple times per week, increasing the need to have additional time in their schedules for both attending class and allowing for study time. The survey asked students how many nights per week they attended classes. The majority of students (60 or 80%) attended classes only one night per week. The ANCOVA for nights per week yielded no significant difference in mean exam values, $F(1, 72) = 1.348, p = 0.249$, partial $\eta^2 = 0.018$.

Some of the students have hometowns more than one hour from the classroom. This commute was considered in the course survey. Although the survey asked students to relate their mileage based upon four categories ranging up to 25 miles from school, many students indicated the miles they were driving. The mileages were entered as the students responded. Forty students in the study group (53.3%) drove less than 10 miles one way to attend class. A group of 26 (35.7%) drove more than 10 but less than 30. The remainder of the students (9 or 12%) drove more than one hour to attend class. Of the nine students driving more than one hour, two students drove more than three hours to attend class. Again, the ANCOVA indicated no
difference between the variables related to driving distance, $F(3, 70) = 0.555$, $p = 0.646$, partial $\eta^2 = 0.023$).

The students who drive may receive some benefit by having another student ride to school with them to decrease stress of driving and also to provide socialization and a potential study partner. Twelve students (16%) had someone who rode to school with them. The ANCOVA indicated no effect on grade differences as a result of having a rider, $F(1, 72) = 0.227$, $p = 0.635$, partial $\eta^2 = 0.003$.

Peer models have been identified as important tools in the process of learning (Bruning et al., 2004). Students in the study group were asked if they had study partners. Nearly half of the students (36 or 48%) did not have study partners. The ANCOVA indicated no significant differences in the mean exam scores for those students with study partners and those without, $F(1, 72) = 0.519$, $p = 0.474$, partial $\eta^2 = 0.007$.

None of the time budget demands revealed significant effects on the exam score differences. However, the students appear to have many demands on their time. Few were able to devote full time to school. Since only three students did not work, the effect of work as compared to not working was not measured and would need further study. Many students had additive effects of work, family obligations and were taking more than one class. This aspect of student life would need more in-depth study than was completed in this research.

No analysis was conducted to determine if combinations of home, work and school constraints had effects on examination scores.

**Research Question Two**

*What importance to practice do nursing graduate students place on science course objectives both knowledge level (undergraduate) and application level (graduate)?*
All of the students enrolled in this pathophysiology course had completed a bachelor’s degree in nursing, which required an undergraduate pathophysiology course. In addition, 12 (16%) of the students had previously completed a master’s degree in nursing and were returning for additional certification or for completion of a family nurse practitioner focus. Previous surveys of nursing students had asked students to rate the value of a science class indicated nursing students place a high value on biology, but believed this content to be too in-depth for their needs (Clarke, 1995; Courtenay, 1991). This research project focused on exploring the content of the course to determine importance of specific content objectives to student clinical practice. Another unique aspect of this study was that experienced nurses who were students would rate these objectives. Previous studies have used surveys focusing on students without clinical expertise or nursing instructors who taught courses related to the content.

Importance for clinical practice was measured through use of a researcher developed survey instrument. The students were asked to rate the course as a whole to the importance to their clinical practice. Additionally, the students were asked to rate individual content objectives with their importance to clinical practice. There were 32 content-level objectives. These were equally split between graduate level objectives and undergraduate level objectives. The students were asked to rate the importance of these objectives to their past clinical practice as registered nurses and the importance of these same objectives to their future clinical practice as nurse practitioners.

The students were asked in the beginning of the survey to rate the overall importance of the course to their clinical practice. The mean score for this value was 4.37 (scale 1 = low to 5 = high, SD = 0.712). However, when asked about specific course content objectives, the mean
value was 2.89 (SD = 0.831) for past importance as registered nurses and 3.84 (SD = 0.746) for future practice as nurse practitioners.

A paired t-test comparison of the mean scores for the overall score importance (M = 4.37) and the average student response for past importance (M = 2.89) indicated the two scores were significantly different (t = 13.57, p < 0.001). A paired t-test comparison of the mean scores for the overall score importance (M = 4.37) and the average student response for future importance (M = 3.84) indicated that these two scores were also significantly different (t = 6.31, p < 0.001). This indicates students have major concerns about some of the content in the course and its importance to practice.

The reliability of this survey instrument was measured through a split-half analysis. The items separated into the graduate and undergraduate objectives. The items were then divided sequentially as they appeared in the survey. The Cronbach’s alpha measures of internal consistency were 0.952 for one half and 0.946 for the second half. The Spearman-Brown coefficient for equal length items was 0.973 and a correlation between forms of 0.948.

Content validity of this survey instrument was determined by using both graduate and undergraduate pathophysiology textbooks. These textbooks were assessed for content level objectives, which were used to develop the objectives in the survey. The leveling of the objectives was achieved by using the texts and also employing the use of Bloom’s Taxonomy to improve leveling (Bloom et al., 1956).

Construct validity of this survey instrument was measured through a factor analysis. The answers were sorted into the past objectives responses and future objectives responses since it was assumed that these would be different measures. Although all factors with eigenvalues
greater than one were considered in the analysis, scree plots were used as the final determination for which factors were included in the matrix.

The factor analysis for all objectives for past importance yielded five factors with eigenvalues greater than 1. The matrix could not be obtained using all five factors and a matrix with three factors eliminated nearly all items as not significantly loading on a single factor so it was decided to use a four-factor elimination in the analysis. The scree plot supported the four-factor matrix.

The factor analysis for all objectives for future importance yielded six factors with eigenvalues greater than one. Two factors were close to one and were numerically closer to additional factors than previous factors. The scree plot suggested four factors should be used in the analysis. The four factors from these two analyses both past and future could be correlated when individual analyses of graduate and undergraduate were separated. Once this was completed the four factors were labeled holistic, cellular, mixed1, and mixed2.

The first major factor contained objectives that dealt with holistic aspects of health care. This includes items where the disease entity related to patient signs or symptoms, related to the function of an organ as a whole, or where measures of care such as electrolyte levels or lab values were used. For example, this group included mechanisms and symptoms of disseminated intravascular coagulation, manifestations of cancer, lab results related to tumors, and differences between syndrome of inappropriate antidiuretic hormone, and diabetes insipidus.

A second major factor contained objectives focused on the cellular or molecular aspects of care without reference to specific organs or tissues within the body. These items included aerobic respiration processes, processes in the formation of proteins, and some genetic mechanisms.
A third factor emerged that included genetic mechanisms related to specific diseases and was labeled mixed1. These included the mechanisms for Down, Kleinfelter, and Cri Du Chat Syndromes. Additionally, this factor included expression of dominant and recessive genes and tissue compatibility in organ transplants.

The fourth mechanism was categorized as mixed2. These objectives referenced cellular communication processes of neurotransmitters, intracellular communication mechanisms, and processes involved in making proteins.

A t-test comparison was conducted to determine if each of the four groupings were identified as significantly different from the others. The holistic factors were found to be significantly different from all of the other three groupings. The holistic factors were rated higher than the cellular factors in every analysis: past and future, graduate and undergraduate (p= 0.000 in each). The holistic factors were significantly higher than the mixed1 and mixed2 factors in every analysis: past and future, graduate and undergraduate (p < 0.001 in each). This would support the notion that students viewed these objectives as more important in their clinical practice.

The cellular groupings, mixed1 groupings and mixed2 groupings were not found to be significantly different from each other using p < 0.008 as a Bonferroni adjustment. The objectives of the mixed1 group were the genetics of Down, Kleinfelter and Cri Du Chat Syndromes, expression of dominant genes, and organ transplantation. The objectives of the mixed2 group were cellular communication and how proteins are made. Therefore all of these groups could be classified using a cellular terminology, such as cellular 1, 2 and 3, rather than the mixed groupings.
The students in this study rated cellular communication and genetic concepts as having little importance to clinical practice. This would imply that a basic understanding of genetics would be helpful but not necessary to their practice. It would be easy to suggest that application level objectives would increase the student’s belief that these concepts are important to clinical practice. However, application level alone may not be sufficient in creating this change in beliefs.

For example, cystic fibrosis (CF) may be more common in areas of the world where diarrheal illnesses such as cholera and typhus are common because of the protection CF offers by not allowing chloride channels to open (Lewis, 2005). This would be application level information.

This survey indicated that it may not be enough to have application level information in a course since several of the objectives were mixed cellular and holistic. The CF example would most likely be labeled by nursing students as interesting or unique, but did not have a direct influence on practice.

However, if they understood how a new class of drugs called “chaperones” could help some patients with CF, they might recognize this as an important process to the care they give the patient. The disease has one basic clinical phenotype, but several genotypes. One genotype responds to typical drug therapy. Another genotype needs the drugs plus the chaperones to respond. This necessitates genetic testing to determine which therapy is most effective.

Chaperones are believed to refold the deformed proteins in some genotypes of CF and convert them to a normal form. The nurse’s role would be to administer the appropriate medications, explain the difference, and to conduct patient teaching about how and why the drugs work or, in some cases, don’t work (Prows & Prows, 2004). Since the nurse needs to teach
patients, there would be an increased need for nurse to understand more about these cellular processes. This would include the basic function of chaperones, the difference between genotype and phenotype, and processes involved in folding proteins. This type of application question would support an increased interest in learning the cellular or molecular aspects of biology.

**Research Question Three**

*How does providing computer-assisted tutorials with feedback change student performance on exams as compared to pencil/paper worksheets without feedback?*

The students were given five paper and pencil exams with worksheets to be used for study guides. The students were not given feedback on the answers. Students could request explanations from each other through a web site posting, but the instructor did not respond or give feedback unless incorrect or misleading answers were posted on the course web site. When this occurred, the instructor gave the correct answer to the question.

The students were given four computer exams and computerized worksheets called tutorials. These tutorials provided feedback to the students immediately following completion of the worksheet. The content was used to identify the four computer examinations that were given in the order: Immune, Fluid & Electrolyte, Renal and Endocrine.

The mean scores on the paper and pencil exams ranged from 77.23 on Exam 1 up to 87.31 on Exam 5. A t-test comparison of these exam scores indicated no significant improvement in scores from Exam 1 to Exam 2 or from either of these to Exam 3. Exam 4 and Exam 5 had significantly higher scores than the first three exam scores, but were not significantly different from each other. This indicates a general increase in scores throughout the
course. It also should be noted that the computer tutorial for the Immune and Fluid Exams occurred between Exam 3 and Exam 4.

The Immune Exam was the first computer examination given and had the highest mean of 90.13. The next highest mean score was the Renal Exam at 87.30, closely followed by the Endocrine Exam at 86.67. The lowest computer mean score was the Fluid & Electrolyte Exam at 84.69. Significant differences were found between the Immune Exam and all other computer exams. No other significant differences were found through t-test analyses.

A comparison of each of the computer examinations with feedback to the paper and pencil examinations without feedback revealed significant differences between all of the computer examinations and paper and pencil examinations 1, 2, and 3 (p < 0.0025 per Bonferroni adjustment, p < 0.001 in all comparisons). The computer examinations were higher than all the first three paper and pencil examinations. The Immune System computer exam was also significantly higher than the paper and pencil Exam 4 (p = 0.001), but not with Exam 5 (p = 0.008). No significant differences were found when comparing Exam 4 and Exam 5 to the Fluid & Electrolyte, Renal or Endocrine computer exams.

Generally, it would appear to indicate the students performed better on the computer examinations overall than the paper and pencil examinations. However, the improvement on the scores for exam 4 and exam 5 would tend to raise some doubt. These examinations were given following initiation of the computer tutorials.

The improvement of mean scores on the later paper and pencil examinations may be attributed to the student’s ability to ascertain how the instructor might ask a question may have influenced this improvement. In addition, the feedback received from the instructor rather from other students on the computer tutorials may have also contributed to the improvement in the
student’s ability to discern how the instructor reasons through a question, thus influencing responses on all examinations.

It was recognized that all of the examinations were different in their content and differences in examination scores may be reflective of different question difficulty. The significant differences in the average scores for the computer examinations (M = 87.23) and the paper and pencil examinations (M = 81.30), (Correlation = -0.744, p < 0.001, t = -9.40, p < 0.000) reduces the potential effect of individual examination content differences.

**Research Question Four**

*Do elaborate feedback answers affect student performance on examinations as compared to verification (correct/incorrect) type feedback?*

Computer environments have been helpful in providing students with immediate feedback for practice problems. Immediate feedback on exam questions is considered to be an important factor in assisting students to form knowledge accurately (Dihoff et al., 2003). This study assessed the effect of different types of feedback.

All students in the study were divided into six groups assigned by a random number generator. The students were assigned to receive elaborate feedback on two examinations and simple feedback on two examinations. This allowed for mixing the students so that different groups of students were used for all four of the computer examinations. The four topics for these computer examinations were Immune, Fluid and Electrolyte, Renal and Endocrine.

Pretesting is a method to establish a baseline in the process of determining growth following an intervention. However, the pretest can alter outcomes since students would be expected to perform better on a second try of the measure (Ary et al., 2002). Therefore, slightly less than half of the questions were pretested. This allows statistical analysis of the pretest vs.
posttest for interventional growth and analysis of pretested vs. nonpretested to determine if the pretest diminished results of the intervention.

Fifteen question numbers were selected at random from each computer examination using a random generator. These questions were used to develop a pretest to determine student prior knowledge. The pretests were administered at least one week prior to any lecture or presentation of any course material on the content. These were administered under a similar testing environment as the examinations. The questions on the pretest were identical to the questions on the examination. The pretest was at least two weeks prior to any examination on the content.

The pretest scores for individual questions were compared to exam scores on the same questions using a t-test. The pretest question scores were significantly different from the question scores on the exam questions for all computer examinations (N = 28, p = 0.000 on all comparisons). The mean scores for the Immune Exam were 42.98 for the pretest and 89.00 for the exam. The mean scores for the Fluid and Electrolyte Exam were 42.26 for the pretest and 84.57 for the exam. The mean scores for the Renal Exam were 43.57 for the pretest and 87.14 for the exam. The mean scores for the Endocrine Exam were 52.14 for the pretest and 88.00 for the exam. The Pretest Score for the renal content was significantly correlated with the Renal Exam Score at 0.391 (p = 0.044). The Pretest Score for the endocrine content was significantly correlated with the Endocrine Exam at 0.450 (p = 0.016). The other pretest scores were not significantly correlated with the exam scores for that content.

The pretested question items were compared to the non-pretested items using a t-test comparison. No significant differences were found on the exam questions between the items that were pretested and the items that were not pretested (t = -0.496, p = 0.621).
The mean scores for the Immune Exam were 89.33 for elaborate feedback and 91.00 for simple feedback \((t = -0.846, p = 0.400)\). The mean scores for the Fluid & Electrolytes Exam were 84.76 for elaborate feedback and 84.61 for simple feedback \((t = -0.053, p = 0.958)\). The mean scores for the Renal Exam were 89.03 for elaborate feedback and 86.00 for simple feedback \((t = 1.347, p = 0.182)\). The mean scores for the Endocrine Exam were 89.03 for elaborate feedback and 86.44 for simple feedback \((t = 0.735, p = 0.464)\).

Since a significant difference was found between the pretested and posttest scores and no significant difference between items pretested and nonpretested items, we conclude that computer tutorials were effective in increasing student performance on an examination without confounding the outcome as a result of prior practice from the pretested items. No statistical significant differences were found between the scores on examinations with simple feedback and the scores where the student had elaborate feedback for all four examinations. This indicates that these graduate students did not benefit from elaborate feedback as compared with simple feedback on the computer examinations.

**Research Question Five**

*Do students retain concepts taught through use of computer tutorials differently than concepts taught using paper and pencil worksheets as measured through a comprehensive final examination?*

Retention of concepts is one measure of self-regulated learning. This study compares scores on a comprehensive final examination to previous examination scores to determine retention of concepts.

The instructor picked questions from each study guide to be used for the Final Exam; 80 questions were selected. All paper and pencil and computer tutorial examinations were used to
develop the Final Exam with the questions derived from those examinations. No questions were repeated verbatim. In the majority of the cases, the root of the question was reworded, but the same concept tested. Often, the answers were reordered or a new foil was added to the distracters. This was designed to assess student learning and retention of concepts rather than student memorization. Students were given the numbers of the questions from the study guides. They had one week to study for the Final Exam that was given in a paper and pencil format.

The questions on the Final Exam were not divided equally among all of the topics in the course. Some content areas had many questions on the final examination while others had only a few. The Endocrine Exam had only three questions for the Final Exam, but the Immune Exam had seven. Students were not given access to previous examinations to study questions nor were they allowed to copy answers from an examination when an examination was reviewed following the initial administration during the semester.

The Final Exam was subjected to two analyses. The first analysis compared the Final Exam total to each of the unit examinations using a t-test comparison. The average on the Final Exam was 79.12. This score was significantly lower than the scores on all of the computer examinations, Exam 4 and Exam 5. No significant difference was detected when comparing the Final Exam score to Exam 1, Exam 2, or Exam 3. The student examination scores had been increasing after exam 3. This would include all of the computer examinations. This was consistent until the Final Exam. The average for the Final Exam declined to a score similar to Exam 1, Exam 2, and Exam 3.

T-test comparisons were made on the content from each of the paper and pencil examinations to the Final Exam questions over the same content. Significant differences occurred between the unit examination score and the Final Exam content for all examinations.
except Exam 5. Exam content from examinations 1, 2, and 3 had improved for the final. Exam 4 content had a significant decline when retested on the Final Exam.

Students demonstrated a decline in scores from the computer examinations to the matching content questions on the Final Exam. The Immune Exam mean declined from 90.13 to 80.33 ($t = 5.02, p < 0.001$) for the immune content on the Final Exam. The Fluid & Electrolyte Exam declined from 84.69 to 62.75 ($t = 6.05, p < 0.001$) for the fluid and electrolyte content on the Final Exam. The Renal Exam declined from 87.31 to 77.00 ($t = 4.47, p < 0.001$) for the renal content on the Final Exam. The Endocrine Exam declined from 87.79 to 67.67 ($t = 7.24, p < 0.001$). This would suggest that retention of these concepts was not as strong as was the retention for the paper and pencil content.

The correlations on the Immune Exam (Correlation = 0.424, $p < 0.001$), the Fluids Exam (Correlation = 0.276, $p = 0.016$), and the Renal Exam (Correlation = 0.282, $p = 0.014$), and the Endocrine Exam (Correlation = 0.268, $p = 0.02$) were significant at $p = 0.05$, but the latter three were not significant after Bonferroni adjustment of $p < 0.0056$.

Only one of the final exam – paper and pencil correlations was found to be significant: Exam 1 correlation was 0.474 ($p < 0.001$). The Exam 5 correlation was 2.94 ($p = 0.010$), but not significant with the Bonferroni adjustment. The correlations for Exam 2 were not significant (Correlation 0.284, $p = 0.014$). Exam 3 correlation was 0.294 ($p = 0.033$). Exam 4 correlation was 0.210 ($p = 0.071$).

The use of reflection in this study did not demonstrate equal retention among all of the examination scores. Further study will need to be completed to better understand the differences in retention. How the students studied for the two types of examinations for the final examination was not addressed. An evaluation of the computer tutorials did not indicate students
went back to the tutorials to study for the computer content on the final examination. The students may have studied differentially for some of the examinations or may have remembered the content differently. All of these areas will need further study. Also, for the reflection phase of the study, students had received feedback on examination scores from the paper and pencil examination questions. They knew where they had made mistakes and may have been able to reflect and address any problems.

The use of self-regulated learning techniques demonstrated improved examination scores until the examination measuring retention. Students seemed enthusiastic about computer examinations and course evaluations indicated that students would like more of these. The scores on the examinations indicated that feedback is critical in student performance in a class. Students need feedback to know what and how to study, especially in a content heavy course. For nursing graduate students, the type of feedback the student received did not appear to affect grade performance. Additional work needs completed to determine how to use the computer-based environment in retention of concepts.