The Relationship Between Progress Style and Exam Performance in Computer-Aided Personalized System of Instruction Courses

by

Crystal R. Springer

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF ARTS

Department of Psychology

University of Manitoba

Winnipeg

Copyright © 2005 by Crystal Springer

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

The Relationship Between Progress Style and Exam Performance in Computer-Aided Personalized System of Instruction Courses

BY

Crystal R. Springer

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of

Manitoba in partial fulfillment of the requirement of the degree

Of

MASTER OF ARTS

Crystal R. Springer © 2006

Permission has been granted to the Library of the University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.

This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.

Abstract

Research on massed vs. distributed practice and on procrastination has turned up a variety of results either supporting or not supporting beneficial effects of evenly distributed progress in a course on exam performance. Archived data from courses using computer-aided personalized system of instruction -- an online, self-paced teaching method -- were used to explore the relationship between the progress styles of students -- specifically, the rate patterns at which they complete unit tests -- and final exam performance. The relationship of peer review points with final exam performance and progress style was also examined. Peer review points are points gained for reviewing other student's unit tests. It was hypothesized that students who distribute completing unit tests evenly would perform better than those whose unit test completion is massed toward the beginning (i.e., early massing) or end (late massing, or procrastination) of the courses. In Study 1 analysis of variance and planned comparisons were used to test the relationship between progress style and final exam performance. There was little or no detectable effect on final exam performance of how progress was distributed provided that all units were completed. In Study 2 results showed only one significant difference in final exam scores, which was between not completing all units and early massed progress. It was also determined that students who earned more peer-review points performed significantly better on the final exam, thus indicating that the lack of difference between progress styles was not due to students who massed their work early in the course being less motivated to perform well on the final exam due to having more peer-review points.

Acknowledgements

I would like to thank Dr. Pear, my advisor, for his time and help. I would also like to thank Dr. Yu and Dr. Kinsner of my examining committee. I would also like to give special thanks to Kirsten Wirth who provided me with the archived data.

Dedication

I would like to dedicate all the work to the people who continued to encourage me through out this process and throughout my education.

Table Of Contents

List of Tables

List of Figures

For centuries lectures have been used for university teaching. In a marked deviation from this approach to education, Keller (1968) designed a self-pacing method of teaching called personalized system of instruction (PSI). This thesis tests the relationship between final exam performances and the manner in which students pace their progress in a computerized version of PSI called computer-aided personalized system of instruction (CAPSI).

Personalized System of Instruction (PSI)

PSI was a system of teaching developed by Keller (1968). It had several principles developed to aid in learning. It was based on behavioural principles of clear specification of the material to be learned, frequent and immediate reinforcement of the material learned, limited punishment and self-pacing (Kinsner & Pear, 1990). Some of the behavioural principles are main fundamentals, which are self-pacing, mastery learning criteria, textual material as the main source of information to be learned, unrestricted opportunities to be re-tested on material, multiple unit testing, and the use of helpers called "proctors" (Sherman & Ruskin, 1978). Proctors are students in a more advanced course who, for part of their course credit, mark and give feedback on unit tests for students in the less advanced course. Mastery learning refers to learning material well enough for it to be considered mastered. For example on a multiple choice test this would require the student to score near 100% to consider mastery in that area of learning. Students receive multiple mastery tests through out the course on each unit and are required to master the unit before progressing to the next unit. The students are allowed to write and re-write these mastery tests as often as they need to master the material. PSI also typically uses reading and handout material to distribute the information. Lectures are used to motivate students rather than the main source of giving information to students. The last fundamental listed and the central part to this research is self-pacing. Self-pacing refers to the student's choice to move through the mastery tests as they prepare for them. Several of these fundamentals are maintained in the CAPSI, but there are also some fundamental differences between CAPSI and PSI. These differences will be discussed later on.

Lecture- Style Teaching

In lecture-style teaching there is no mastery requirement; usually students are tested a few times during the course, and the passing criterion is typically set low (e.g. 50% or 60%). As a result, amount of feedback students receive from the instructor can be very low. There are also no peer-reviewers in regular lecture-style teaching. Although students may comment on each other's arguments and ideas during class discussions and students may sometimes mark each other's quizzes, feedback from peers is usually low.

Research on PSI

A number of studies have examined the overall effectiveness of PSI and conducted analyses of its components (e.g., Canelos & Ozbeki, 1983; Gray, Buerkel-Rothfus & Yerby, 1986; Kulik, Kulik, & Cohen, 1979; Murphy, McMichael & Cariello, 1977; Pickthorne & Wheldall, 1982; Schmitt, 1998). Regarding overall effectiveness, a meta-analysis by Kulik et al. (1979) examined 61 studies on PSI. A total of 57 of these found final exam scores better for PSI-taught courses than for conventional teaching styles. The average final exam score for PSI courses was 73.6%, whereas conventional teaching-style courses produced an average final exam score of 65.9%. This meta-analysis clearly supports the greater effectiveness of PSI over lecture-style teaching methods. Study time appeared not to differ between the two types of teaching. This disconfirms an earlier hypothesis by Kulik, Kulik, and Carmichael (1974) that students in PSI courses do better than students in lecture-style courses because the students in PSI courses spend more time on their course work.

There have been other studies on the overall effectiveness of PSI since the meta-analysis by Kulik et al. (1979). For example, Canelos and Ozbeki (1983) compared problem solving ability between a PSI group and a lecture/recitation group. The results again supported PSI as a superior method of instruction. PSI students improved their problems solving skills and showed a significantly greater ability to answer high-difficulty questions. Gray et al. (1986) compared students in PSI course with those in a lecture/recitation format in a communication course. Once again the results support PSI in that the PSI students had less speech anxiety, greater course satisfaction, and higher final exam scores. Pickthorne and Wheldall (1982) taught physics to engineering students using a PSI method compared to a lecture format method of instruction. The findings again supported PSI as a more effective method of instruction. Students in the PSI group averaged 10% higher on the final exam than the lecture group did.

Regarding research on the components of PSI, Murphy et al. (1977) examined final exam performance as a function of unit quiz retakes and found no correlation between number of unit quizzes students took and final exam performance. These results contradict previous results by Whitehurst and Madigan (1975) who found that number of unit quiz retakes improved retention of material and therefore lead to high final exam scores. Either result would have some bearing on the effects of procrastination, which is the focus of this thesis. On the one hand, it would seem that individuals who do not procrastinate would have more time to retake quizzes. On the other hand, if individuals have more time to prepare for their quizzes they will not have to retake them as often. Students who did not procrastinate would have more time to prepare for the unit quizzes so that they do not need to retake them—possibly explaining the lack of significant findings by Murphy et al.

According to Brothen and Wambach (2001) students who use what is called the "prepare-gather feedback-restudy" method before taking computerized multiple-choice quizzes scored higher on the unit quizzes, had lower quiz times and fewer quizzes taken than students who used the quizzes simply to learn the material. Brothen and Wambach maintain that the prepare-gather feedback-restudy is the most appropriate method for mastery tests. Preparing refers to a student reading through and studying the material for the mastery test first. If they are not successful, "gathering feedback" from the student proctors is the next step. The final step is "restudying" all the material before re-testing.

In a meta-analysis Kulik, Kulick and Bangert-Drowns (1990) examined effects of mastery learning—which, as indicated, is a component of PSI. The findings indicated positive effects on exam performance for students at multiple age and education levels. They also noted that the weakest students showed the strongest effects.

Schmitt (1998) formulated a relationship that helps to explain how self-pacing and memory retention affect performance in PSI-taught courses. He suggested that there is a logarithmic function that describes the amount remembered plotted over time. According to this function, as time increases between study sessions, material learned also increases—but with a diminishing amount of increase. He further suggested that consistent and increased connection with the material used is important in retention. He also recognized organization of information as a key component of learning and memory (Reisberg, 1997). The relationship between self-pacing rates and performance outcome will be explored further later.

Computer-Based Teaching

Studies indicate that computers' can be an effective learning tool when used appropriately. Bangert-Drowns, Kulik, and Kulik (1985) in a meta-analysis found computer-assisted instruction and computer-managed instruction increased students' exam scores. Thus, CAPSI should be effective because its use of computers in a manner designed to maximize learning as well as the fact that it is a variant of PSI.

CAPSI

The University of Manitoba offers several undergraduate psychology courses that use CAPSI. These courses do not have assigned class times and students work at their own pace (within course limits). CAPSI, like PSI, follow the same behavioural principles and incorporates similar fundamentals, but into a computer context (Kinsner & Pear, 1990; 1988; Pear & Novak, 1996). As a variant of PSI, CAPSI should be a highly effective teaching method, although it has not yet been directly compared to traditional lecture methods. CAPSI is similar to standard PSI in several ways (Pear & Novak, 1996): the material is divided into a number of study units; there is a mastery test on each study unit; the answers to the questions on each unit test must be entirely correct (by being both accurate and complete) before the student is allowed to write a test on the next unit; and there is no specified limit on the number of times a student may be re-tested on a given unit. In lieu of proctors as the primary source of student help, CAPSI-taught courses use peer reviewers, who are students in the course who are chosen to review or mark classmates' unit tests. To be eligible to be a peer reviewer on a unit, a student must have demonstrated mastery on that unit. CAPSI-taught courses use written-answer questions and students demonstrate mastery by not omitting or misstating factual information relevant to answering the questions on unit tests. Each unit test is marked by two peer reviewers who have passed that unit or by a teaching assistant or the instructor when eligible peer reviewers are unavailable. As a measure of quality control, when two peer reviewers mark a test both must rate the test a pass or it is considered a "restudy." There is an appeal process whereby appeals on restudies are directed to the instructor. Peer reviews earn points, called peer review points, towards their final grade for performing peer reviewer functions. Each time they review another students unit mastery tests they receive credit up to a maximum amount. CAPSI-taught courses typically also have final exams and one or two midterms given on specified dates and times, which somewhat limits the self-pacing aspect of the system.

The large fundamental difference between PSI and CAPSI is the use of the computer medium. In PSI there are face-to-face test taking times with the proctor. However, with CAPSI students are able to write their test at anytime over the computer and all feedback is through the computer, creating a different system of interaction between peer reviewers and students. The use of computers allows more flexibility in completing tests and communicating with other students and instructors. Face-to-face contact can be arranged through regularly scheduled office hours, but is not required. The use of computers also allows for precise tracking of each student's unit test taking and their peer review points, as well as providing structure for these activities of the course. This use of computer-mediated communication is unique to the CAPSI program and differentiates it from PSI.

Self-Pacing and CAPSI

The main focus of this paper is the self-pacing aspect of CAPSI. In a self-paced course, students may work at either a massed progress or distributed progress pace (recognizing, of course, that there these are actually the two end points on a continuum). Massed progress is defined as compressing work completion (study time, assignments) into a shorter amount of time than is available to it. This research is concerned with two types of massed progress: (1) late massed progress, defined as completely massing all work required near the deadline time; and (2) early massed progress, defined as completely massing all work required at the beginning of the specified time interval given to complete the task. The former is also called "procrastination."

An example of late massed progress would be writing a paper the night before it is due, even though the paper was assigned three weeks ago. An example of early massed progress is writing that same paper in one night, the same day that it was assigned and three weeks before it is due. Distributed progress is defined as the even allotment of progress or work completion from the time of the assignment until the designated time limit for its completion has expired.

The concepts of massed and distributed progress are related to the concepts of massed and distributed practice. In the former, new material is added as previous material is learned, whereas in the latter the specific task does not change. An example of distributed practice is one hour of baseball practice twice a week throughout the baseball season. As example of (late) massed practice is slacking off most of the season and then practicing continuously for many hours just before the final game. Many researchers have studied the effects of massed and distributed practice (Bahrick & Phelps, 1987; Dempster, 1988; Metalis, 1985; Nunnally, Duchnowski & Knott, 1967; Rider & Abdulahad, 1991; Zimmer & Hocevar, 1994). Their research will be discussed later.

The question addressed by this study is what is the relationship between type of self-pacing progress style and final exam performance within a CAPSI-taught course? In particular, is allowing students to choose their own pace without adverse final exam consequences? This research does not directly address this question; instead, it focuses on a preliminary question: which style of progress that students choose in a self-paced procedure appears to be most effective? If one appears to be more effective than another, this would suggest that students should be encouraged, or perhaps even required, to adopt that style (e.g., by imposing a time limit on completion of each unit).

Practice and Progress Style Research

As mentioned there are three types of progress styles: early massed progress, late massed progress (procrastination), and distributed progress. As already indicated, research has directly compared massed versus distributed practice (practice referring to repeating the same material to be learned as compared to progress where there is less repetition and new material is being added) (Cook, 1934; Bahrick & Phelps, 1987; Dempster, 1987; Metalis, 1987; Nunally, Duchnowski & Knott, 1967; Rider & Abdulahad, 1991; Webb, 1933). Some research has also examined the effects of distributed progress versus procrastination (Tice & Baumeister, 1997; Zimmer &

Hocevar, 1994). Research in this area appears to have generated mixed results, indicating support both for and against distributed practice and distributed progress.

Massed versus distributed practice research. Massed versus distributed practice in learning has a relatively long history of research (Cook, 1934; Webb, 1933). In the 1930s, researchers compared massed to distributed practice for physical mastery of a skill (Webb, 1933) and for mental mastery tasks (Cook, 1934). Webb used a pursuitmeter, which is a device that promotes the learning of physical coordination and reaction, requiring subjects to maintain contact of a stylus with a moving bead. The four conditions used were twelve tries in a row, one try daily for 12 days, one try per week for 12 weeks, and 3 tries per week for four weeks. He found no significant difference between massed and distributed practice. However, to confuse matters, Webb's findings suggest that starting out with the most spaced practice schedule and moving to a less spaced schedule appears to be most effective.

Cook (1934) did not obtain the same results as Webb (1933). In Cook's research subjects were required to learn to solve a puzzle as fast as they could. There were two experiments; each experiment compared massed and distributed practice at solving a puzzle. However, there were different puzzles for the two experiments with varying degrees of difficulty. In the first experiment massed practice was better than distributed practice with regard to speed at solving the puzzle. In the second experiment massed practice was again better than distributed practice with regard to speed of solving the puzzle, until after 9 to 11 trials. After that there was no further advantage for massed practice. In addition, although for immediate recall massed practice was better, for delayed recall distributed practice was better. This finding is consistent with later suggestions by Dempster (1988). Thus Cook found an exception to the superiority of distributed practice; namely, on tests of immediate recall. Both of the above experiments therefore seem inconclusive in determining which practice style is more effective. However, this may have been due to the type of task required.

Metalis (1985) examined massed and distributed practice effects on video-game skill acquisition. Massed practice consisted of successive game playing, whereas with distributed practice there was a two-minute interval between games played. The results showed that both groups made improvements; however, the distributed-practice group showed greater improvement across nine games.

In an experiment focused on learning fine and gross motor skills, subjects showed no significant difference in learning using massed or distributed practice (Rider & Abdulahad, 1991). Subjects, who were educable mentally handicap students, were assigned to either a massed or a distributed practice schedule. The massed schedule consisted of 15 30s trials, with 5s rests between trials. The distributed schedule consisted of ten trials with 30s rest on one day and five more trials 24 hours later all with 30s rests. They found no significant differences in learning between groups with regard to practice schedule (Rider & Abdulahad, 1991). The authors suggest that after a certain amount of practice a level of proficiency may develop regardless of practice type. This research indicates that practice schedule may not be an important aspect of learning a new skill. Actually any pace could be equally effective, unlike what is indicated by the research previously discussed (Cook, 1934; Metalis, 1985; Webb, 1933) which suggested a specific practice schedule for the best results.

Nunnally et al. (1967) found no significant difference between massed practice and distributed practice groups. They used four different practice schedules: all 90 trials in one session, 30 trials per session for three consecutive days, 30 trials per session with sessions two days apart, and 30 trials per session with one session per week. They found no significant difference in final performance between the groups, again suggesting that pace does not affect final performance.

Dempster (1987) conducted research on distributed progress in learning vocabulary words. The vocabulary words were presented with or without interpolated words. The former condition was considered distributed practice, while the latter condition was massed progress. These practice styles were contrasted with two different encoding styles. In two experiments Dempster (1987) found that distributed practice resulted in more correct definitions given, regardless of encoding method. This supports the view that distributed practice is more effective in learning. One strong difference in this research and the research conducted in this thesis is the time periods involved. In the research by Dempster and other research on massed vs. distributed practice or progress, time between trials was measured in minutes while in this thesis time between tests was measured in days.

However, Bahrick and Phelps (1987) found some support for the view that very long time intervals between learning sessions increased retention. They tested subjects who in a study by Bahrick (1979) eight years previously had learned English-Spanish word pairs under different interval schedules. The three practice schedules had been zero time between learning sessions, one day between learning sessions, and 30 days between learning sessions. Bahrick's original findings were that learning was best with the shorter inter-session intervals, but that after seven sessions, regardless of interval, retention was almost perfect. Therefore, the results first supported massed practice. Bahrick and Phelps found that after eight years, however, recall was 2.5 times higher for subjects who were in the 30-day inter-session condition than for subjects in the zero inter-session condition. This supports the view that distributed practice is more effective for long-term retention of material.

Dempster (1988) claimed that research on improved performance due to distributed practice or distributed progress has not been applied to the real world. In other words – according to Dempster – distributed practice is more effective than massed practice, yet distributed practice is still not consistently used as a method of learning new information.

Dempster (1988) went on to claim that distributed practice has consistently proved to be better than massed practice with few exceptions. These exceptions are tests of immediate recall, tests of simple isolated skills, and for children less than seven years of age. Another exception may be that massed practice is more effective if paraphrase versions rather than verbatim versions are required. A final exception or limitation to the superiority of distributed practice is that beyond a certain spacing, further increases between practice times are not always associated with further increases in learning (as indicated by the logarithmic relation formulated by Schmitt [1998], mentioned previously). These exceptions may explain the greatly varying results in distributed versus massed practice research.

Massed versus distributed progress research. Most of the research that has been presented on massed and distributed practice does not involve academic skill or information acquisition. The research that examines academic skill or continually has new information to learn generally uses distributed progress or massed progress schedules. Zimmer and Hocevar (1994) tested massed and distributed progress effects on test performance and anxiety. The distributed condition consisted of ten 10-item exams spaced one week apart over ten weeks; the massed condition consisted of one 100-item final exam at the end of the course. The results showed that the spaced exams (distributed progress) had a greater positive effect on classroom performance. This supports the effectiveness of the spacing effect or distributed progress throughout the course work.

Research on the effects of procrastination/late massed progress. As defined earlier procrastination is a style of self-pacing—i.e., late massed progress. Many researchers (e.g., Rothblum, Solomon & Murakami, 1986; Solomon & Rothblum, 1984; Steel et al., 2001; Tice & Baumeister, 1997) have studied student's procrastination in university, measuring its effect on stress, anxiety, mood, and performance. According to Soloman and Rothblum (1984) over 50% of students procrastinate. This provides a great opportunity to study the effectiveness of this form of self-pacing.

Some procrastinators may claim that they work more effectively at the last minute, pooling all their resources to focus on the task. Or they may claim that they work best under the pressure of a deadline. Some research suggests otherwise (Tice & Baumeister, 1997), while, as noted earlier, some research suggests that there is no detrimental effect from procrastination or late massed practice/progress (Nunnally et al., 1967; Rider & Abdulahad, 1991). Tice and Baumeister (1997) consistently found a negative correlation between procrastination and performance, defined as grades on term papers and exams. The one benefit noted was that procrastinators tend to delay feeling stressed, but had higher stress levels later in the semester (Tice & Baumeister, 1997). This research indicates that distributed progress is the more effective than late massed progress.

The research on progress styles appears to support distributed progress for the best performance results. However, research on massed and distributed practice appears to be mixed. Some research suggests distributed practice or progress is more effective (Bahrick & Phelps, 1987; Dempster, 1987; Metalis, 1985; Steel et al., 2001; Tice & Baumeister, 1997; Zimmer & Hocevar, 1994). Other research is neutral or inconclusive about the effectiveness of spacing practice, suggesting that spacing may not play an important role in final performance (Cook, 1934; Nunally et al., 1967; Rider & Abdulahad, 1991; Webb, 1933). There may be an explanation for why the spacing effect did not work in those experiments listed. As mentioned by Dempster (1988) there may be exceptions to when distributed practice or progress is more effective than massed practice or progress.

Research on the Effects of Self-Pacing.

Research on the effects of self-pacing also has produced mixed results. Some studies indicate that allowing students to choose their own pace has beneficial effects (Reiser, 1984; Santogrossi & Roberts, 1978; Wesp & Ford, 1982), whereas other studies suggest that allowing students to choose their own pace has detrimental effects (Ainsworth, 1979; Brooke & Ruthven, 1984; Hobbs, 1981; Steel, Brothen & Wambach, 2001; Wesp, 1986).

Studies indicating that self-pacing may be harmful. Steel et al. (2001) used a computerized introductory psychology course taught with a computerized version of PSI.

In this research, PSI consisted of students working independently at their own pace to complete computerized vocabulary quizzes. Prior to taking the quizzes students studied by reading the text and completing study questions and practice tests. Unlike PSI as defined in this paper, the proctors provided assistance but did not provide feedback for the computerized tests. The researchers found that procrastination was negatively correlated with exercises completed (-.94) and final exam grade (-.38). Negative correlations were also found for self-reported procrastination and exercise completion, but were not as strong. One interesting correlation is the relatively weak correlation of final exam grade with procrastination. Steel et al. noted that some students were able to "catch up". That is, students who procrastinated but still managed to complete all the assignments tended to perform as well on the final exam as students who did not procrastinate, thus contradicting the view that procrastination necessarily lowers final exam performance.

Other research suggests that self-pacing is harmful to the student, especially if the student is prone to procrastinating (Ainsworth, 1979). Hobbs (1981) also found that self-pacing for students' leads to inferior performance compared to instructor pacing.

Brooke and Ruthven (1984) found that deadline contracts for mastering unit material increased final exam scores. They set up short and long term contracts for assignment deadlines and compared these students performance against performance of students without contracts who were allowed to complete the assignments at their-own-pace. The students in the contract groups had higher final exam scores than those in the self-paced group. Wesp (1986) also found support for controlling the pace of the course material. He set a requirement of daily quizzes or self-initiated quizzes. The results showed that students who had required daily quizzes moved through the course more quickly and earned higher grades than the students in the self-initiated-quizzes group.

Studies indicating that self-pacing may not be harmful. Santogrossi and Roberts (1978) showed that instructional methods directing pacing in a course did not affect grade performance. Their findings indicate that directing the pace of instruction did not mask the effects of prior aptitudes, preparation, and intelligence. Student's scores were not affected by self-paced progress. This also indicates that procrastination may not be as negatively correlated with learning as might be expected.

Wesp and Ford (1982) also explored progress strategies in PSI courses. They explored little, moderate, and extreme degrees of student choice in pacing. Somewhat contrary to the previously presented research they found that some pacing optimized performance but suggested that self-pacing with some limitations may be an effective strategy.

Reiser (1984) studied a penalization strategy to reduce procrastination by students in a PSI formatted course. Students in this course were penalized for failure to follow an assignment schedule. Students performed equally on the final exam whether they were on a penalization schedule for their assignments or not. A secondary finding was that penalized subjects moved more rapidly through the course. Even though these subjects moved more rapidly and on schedule through the course material, they did not perform better on the final exam. This suggests that self-pacing is not harmful to performance and allowing students to move through the material at their own chosen rate has no negative effect on outcome performance.

Thus, research on self-pacing in PSI seems to be inconclusive. Some research suggested that allowing self-pacing does not affect overall grade performance (Reiser, 1984; Santogrossi & Roberts, 1978; Wesp & Ford, 1982). Other research suggested that scheduling the pace of the course material leads to better grades and faster movement through the courses (Ainsworth, 1979; Brooke & Ruthven, 1984; Hobbs, 1981; Lamwers & Jazwinski, 1989; Ross & McBean, 1995; Wesp, 1986).

Research on effects of self-pacing on movement through a course. Some research has examined ways to modify self-pacing rates without regard to the effect on examination performance. Ross and McBean (1995) compared multiple deadlines throughout the semester to a single deadline format for all assignments in a version of PSI. The results showed that test taking was steadier with fewer pauses when there were variable multiple deadlines. Thus having deadlines for the mastery quizzes keeps students working steadily and stops them from procrastinating.

Lamwers and Jazwinski (1989) observed that students in PSI courses who were allowed to self-pace tended to procrastinate. These researchers explored strategies to decrease the amount of procrastination and found that students who were given credit for completing assignments before a certain time progressed through the course faster and were more likely to complete the course. This supports the view that complete self-pacing is not the ideal strategy for learning in PSI courses—at least those that have course deadlines.

Statement of the Problem

The purpose of this study was to examine procrastination in CAPSI-taught courses. Because there is a limitation on amount learned across time, as suggested by

Schmitt (1988), it is hypothesized that distributed progress would be more effective because the student would start a new learning curve each time they had a study session. These students would be in the peak learning time more often than students who massed their progress. If the learning curve were a logarithmic function of material learned over time, most material learned would be at the beginning of the study session. So, for example, if a student studied two hours a night for four nights, he or she would spend more time in the beginning of the learning curve and make greater gains than a student who studied eight hours in one evening. This is because the latter student would spend more time in a flatter area of the learning curve (or area where less material is learned per unit time).

A secondary question is the effect of peer reviewer points on final exam performance, because students who massed their units early would have more opportunities to peer review than would students who did not mass their units early. Having a greater number of peer-review points could act as a negative incentive (i.e., a disestablishing operation) to attempt to do well on the final. However, it seems that students who did more peer reviews would tend to do well on the final by virtue of having reviewed the material more often. If there is a negative correlation between number of peer review points and final exam performance, this could suggest a potential confound between the effects of early massing and that of number of peer review points on final exam performance.

Method

Subjects and Data

Data from four undergraduate psychology courses (17.244, 17.245, 17.247 and 17.252) taught using CAPSI at the University of Manitoba during the fall semester of 2000 and 2001 were obtained from archived files. Course 17.244 (Behaviour Modification Principles) covered fundamentals principles and procedures of behaviour modification. Course 17.245 (Behaviour Modification Applications) covered designing, implementing, and evaluating applications of behaviour modification. Course 17.247 (Learning Foundations of Psychology) covered basic conditioning principles and their relation to complex animal and human behaviour. Course 17.252 (Orientations to Psychological Systems) compared and contrasted humanistic, psychodynamic, cognitive, and behavioural psychology with regard to fundamental issues in psychology. The total number of students who wrote the final exam in all the courses was 96, and only the data from those students were included.

The material for each course was divided into 10 units, with 98 days of the course available to complete all units. Students proceeded through the units at their own pace, according to the CAPSI procedure. In this procedure, students write brief essay-type tests based on study questions. An online program provides unit tests to students who request them, and submits completed tests to peer reviewers who have completed the unit being marked. Students must have demonstrated mastery on a given unit in order to be a peer reviewer for that unit. (See Appendix A for further details regarding the CAPSI program). In addition to the unit tests, there were two midterm exams and one final exam in each course. All examinations were based on the study questions in the course.

Study 1

The data analyzed in this study consisted of the number of days in the course each student completed each unit, and each student's midterm and final exam scores. The data of students who did not finish all ten units were excluded; leaving a total of 58 students that completed all the units and the final exam.

Method of analysis

A procedure for finding the area under a curve was used to obtain what is called a *Progress Score* for each student, prior to each exam. The curve from which this area was obtained was a graph of the student's progress through the units, with the horizontal axis representing days and the vertical axis representing number of units completed.

For each student the total number of units completed at the end of each day was recorded in a table. These numbers were then summed over the 98 days. This sum is called the progress score for each student. To take an extreme and completely improbable example, if the student completed all ten units in the first day of the course, the student's progress score would be 10 added 98 times, or 980. At the opposite extreme, if the student completed all 10 units on the last day, the progress score would be 0 added 97 times plus 10, or in other words, 10.

Thus, progress scores for Study 1 fall between 10 and 980, and indicate all three types of progress as defined earlier. Higher scores indicated early massed progress, lower scores indicate late massed progress, and intermediate scores indicate distributed progress. See Table 1 for examples of these three different types of progress.

Table 1. Example of computing distributed, massed/late and massed/early progress scores.

	Day														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Area
Distributed	0	1	2	3	3	4	5	6	7	7	8	9	10	10	75
Mass/late	0	0	0	0	0	0	2	4	4	5	6	7	7	10	45
Mass/early	3	5	6	8	9	9	10	10	10	10	10	10	10	10	120

Progress scores were divided into the three comparison groups: early massed progress, late massed progress, and distributed progress. Of course, there are no clear dividing lines between these three types of progress; therefore, the dividing lines were defined somewhat arbitrarily, and by approximate percentile rank. Early massed progress for the final exam was defined as scores above 507, for the first midterm it was defined as scores above 61, and for the second midterm it was defined as scores above 248. Distributed progress for the final exam was defined as scores between 393 and 507 (inclusive), for the first midterm it was defined as scores between 36 and 61 (inclusive), and for the second midterm it was defined as scores between 165 and 248 (inclusive). Late massed progress for the final exam was defined as scores below 393, for the first midterm it was defined as scores below 36, and for the second midterm it was defined as scores below 165.

To ensure objectivity in the dependent variable (performance on the final exam and the two midterm exams), the examinations were re-scored with clear operational definitions, and inter-observer reliability measures were taken on scoring the exams for each course. Crone-Todd, Pear, and Read (2000) produced the operational definitions for her research of higher-order thinking in post-secondary education. Each set of exams used three scorers, the scores were teaching assistants from the course and graduate research assistants. The level of agreement of the exam scores was evaluated with the Kappa statistic. This statistic corrects for chance agreement between scorers. A minimum Kappa of .80 was required. For each course the minimum requirement of 0.80 was met, with the range being from 0.936 to 0.836.

The exam scores were converted to *z* scores to ensure comparability of each of these sets of scores across the courses. Tests of normality were conducted for the distribution of the exam scores. This was done with the Shapiro-Wilkes test of normality and visual analysis of stem and leaf plots, box plots, and scatter plots. The Shapiro-Wilkes test indicated that the final exam z scores were within normal range ($p = .173$). The plots for the final exam scores also indicated normal distribution of scores. (For plots of normality for the final exam see Appendix B). The Shapiro-Wilkes test of normality did not indicate normal distribution of scores for midterm exam 1 ($p = .000$) or midterm exam 2 ($p = .001$). The midterm exam scores were skewed, which was corrected by squaring the exam scores, this was done for both midterm exam z scores. Figure 1 shows the skewed midterm exam 1 before correction and Figure 2 shows midterm exam 1 after correction.

Frequency	Stem &	Leaf
3.00	-2	023
3.00	-1	778
3.00	-1	034
6.00	-0	555568
6.00	-0	233334
18.00	0	000001224444444444
18.00	0	567888888999999999
.00	1	
1.0	1	6

Figure 1. Stem-and-Leaf Graph of Midterm Exam 1 before crrection for Skewness.

Frequency	Stem &	Leaf
6.00	-1	567788
3.00	-1	134
9.00	-0	777777799
14.00	-0	01111111111111
2.00	0	22
19.00	0	5777777777777777777
1.00	1	3
1.00	1	6
3.00	2	000

Figure 2. Stem-and-Leaf Graph of Midterm Exam 1 after correction for Skewness (For more plots of normality for midterm exams see Appendix C).

Each student was also put into progress groups for each midterm exam and the final exam. The analysis comparing the exam performances of the progress groups was done using the Dunnett T3 method of multiple comparisons. The Dunnett T3 method is designed for multiple planned comparisons of one mean to other means when there are unequal group sizes and unequal variances. In this study, the distributed progress group was the designated mean to which the other means were compared. That is, the distributed progress group mean was compared to the early massed progress group exam score means and the late massed progress group exam score means. The distributed progress group mean on each exam was also compared to the overall massed group progress mean on that exam to detect any differences in exam performance between distributed progress and massed progress in general.

Results

Table 2 presents the overall analysis of variance (ANOVA) results for final exam scores in the Early-Massed, Late-Massed, and Distributed Progress Groups. There were 58 subjects in the analysis. Note that overall differences were not significant at the p = .05 level.

Table 2. ANOVA for the final exam scores of the Early-Massed, Late-Massed, and Distributed-Progress Groups

	Sum of Squares	Df	Mean Square	F	Significance
Between Groups	3.244	2	1.622	1.897	.160
Within Groups	47.033	55	.855		
Total	50.277	57			

Table 3 shows the result of the Dunnett T3 comparison of the Early- and Late-Massed Groups with the Distributed-Progress Group on the final exam. Note that although the mean of the Early-Massed Group was higher than the mean of the Distributed Progress Group (mean $z_{\text{Early-Massed}}$ - mean $z_{\text{Distributed-Progress}}$ = .51), the difference was not significant at the p = .05 level. The difference between the mean of the Late-Massed Group and Distributed Progress was even smaller (mean $z_{\text{Late-Massed}}$ - mean $z_{\text{Distributed-Progress}}$ = .08).

Table 3. Dunnett T3 Comparison of Early-Massed and Late-Massed Groups with the Distributed Progress Group

Progress Style	Mean Difference	Standard Error	Significance
Early Massed	.5129814	.30043140	.257
Late Massed	-.0843709	.30633303	.990

Tables 4 and 5 present the overall analysis of variance (ANOVA) results for the scores on Midterms 1 and 2, respectively, in the Early-Massed, Late-Massed, and Distributed Progress Groups. Note that overall differences were not significant at the p = .05 level.

Table 4. ANOVA for midterm 1 scores of the Early-Massed, Late-Massed, and Distributed-Progress Groups

	Sum of Squares	Df	Mean Square	F	Significance
Between Groups	1.158	2	.579	.639	.532
Within Groups	49.818	55	.906		
Total	50.976	57			

Table 5. ANOVA for midterm 2 scores of the Early-Massed, Late-Massed, and Distributed-Progress Groups.

	Sum of Squares	Df	Mean Square	F	Significance
Between Groups	4.676	2	2.338	2.724	.074
Within Groups	47.209	55	.858		
Total	51.884	57			

Table 6 shows the means in z scores for all three groups on all three exams. Each *z* score for a particular exam was calculated from the mean and standard deviation for that exam. Note that the Early-Massed Progress Group has the highest mean for all three exams.

Table 6. Final exam and midterm exams mean *z* scores (based on the three main groups).

Progress Style	Final Exam Mean	Midterm Exam 1 Mean	Midterm Exam 2 Mean
Late Massed	-.1721553	-.1149720	-.4334641
Distributed	-.2565263	-.0794608	-.0181318
Early Massed	.2564551	.1907807	.2838042

Discussion

The data suggests that students who early mass may have higher final exam scores that students who late mass or distribute progress. However, if the difference exists it was not large enough to reach significance despite the relatively large sample size. The statistical analysis in Study 1 did not indicate any significant difference between any of the progress style groups on the final exam. There was no significant difference between groups in midterm exam 1 or 2 either. This may support the findings of Steel et al. (2001) indicating that students can "catch up" to other students as long as they finish all the units. It appears that allowing students to choose their own pace does not substantially harm their exam results.

A second point of interest is that the groups were chosen by their percentile rank. If the groups were in a controlled rather than applied research setting the early massed group may have more resembled a distributed progress, therefore supporting the idea of the distributed progress being the more effective method of progress. As noted in the methods the maximum progress score for the final exam was 980 and the minimum was

10. However, the early massed progress minimum was at 508, which is almost the mid-point of this range, rather than the 66th percentile of this range. The distributed progress range for the final exam was from 393 to 507, a much smaller range than a third of the entire range. The range for the late massed progress was also not a third of the entire range. The actual range of progress scores were 508 to 788 for early massed, 507 to 393 for distributed and 169 to 392 for late massed progress. Neither early or late massed progress groups were particularly close to the end points of the experimental range. Again the early massed group had a much larger range in progress scores. This may have affected the results that insignificantly showed that early massed progress had the highest average final exam score. This may also be true for the midterms.

Study 2

Since there was no significant difference between late massed progress and distributed progress and no significance indicated between any other groups, another Dunnett planned comparison was conducted. Research by Steel et al. (2001), using a different approach, indicates that students may be able to "catch-up" to other students in their learning of the course material by massing late progress. That is, if these late-progress students complete all the units in the short time near the end of the course, there may be no significant negative effects on final exam scores. The question arises, however, as to whether there is a negative effect if students do not "catch up" to other students by completing all the units by the end of the course.

Study 2 used the same method as Study 1 used except that students who did not complete all units were included. This increased the number of subjects to 96. The fourth group, "Incomplete," was added so that a comparison could be made between students in which all the units were not completed and the three progress groups' final- and midterm-exam scores. The absence of a difference between these groups would indicate that students do not need to complete all study units to perform successfully on the midterm and final exams.

Study 2 also looked at the correlation between number of peer review points on final exam performance. If this correlation were negative, it could indicate that students who obtained a large number of peer-review points were less motivated to do well on the final exam. Since students who finish units early would have more opportunity to earn peer-review points, this could potentially confound the results by introducing a bias against early massing.

The progress group peer review points were compared to examine whether there is a significant difference between any of the groups in proctor points received. Again a Dunnett T3 test was used with the incomplete progress group as the comparison mean.

Results

Figure 3 shows box plots that were produced to provide a visual impression of the distributions of the four groups. The box plots show that although there was a great deal of overlap between the groups, the Incomplete Group had the lowest mean z score and the Early-Massed Group had the highest. Note also that the massed groups—especially the Late Massed Group—had less variability than the other groups.

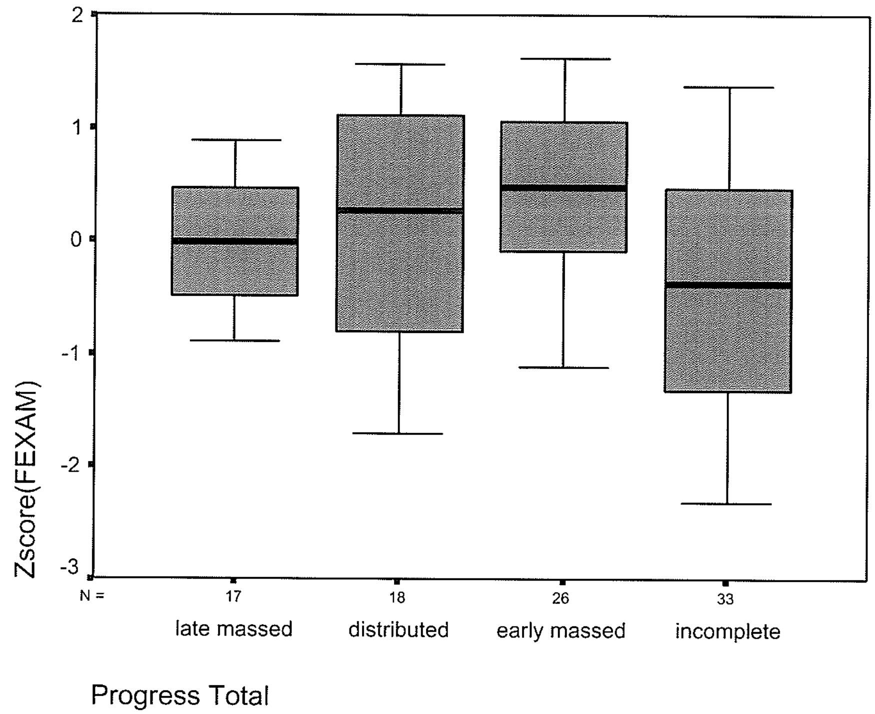


Figure 3. Boxplots of the distribution of each progress group

Progress through the course was examined by visual inspection. Progress trends around the two-midterm times were examined. On Figure 4 the midterms at the 4.5 and 9. Students that tended to late mass progress may have practiced this up to each midterm time. As shown in Figure 4 the progress lines for incomplete and late massed progress are below the early massed progress and distributed progress group. The incomplete and late massed progress group tend to have two exponential growths in them at the time of the midterms, this would tend to support the idea that students in those two groups tended to mass progress at times during the course.

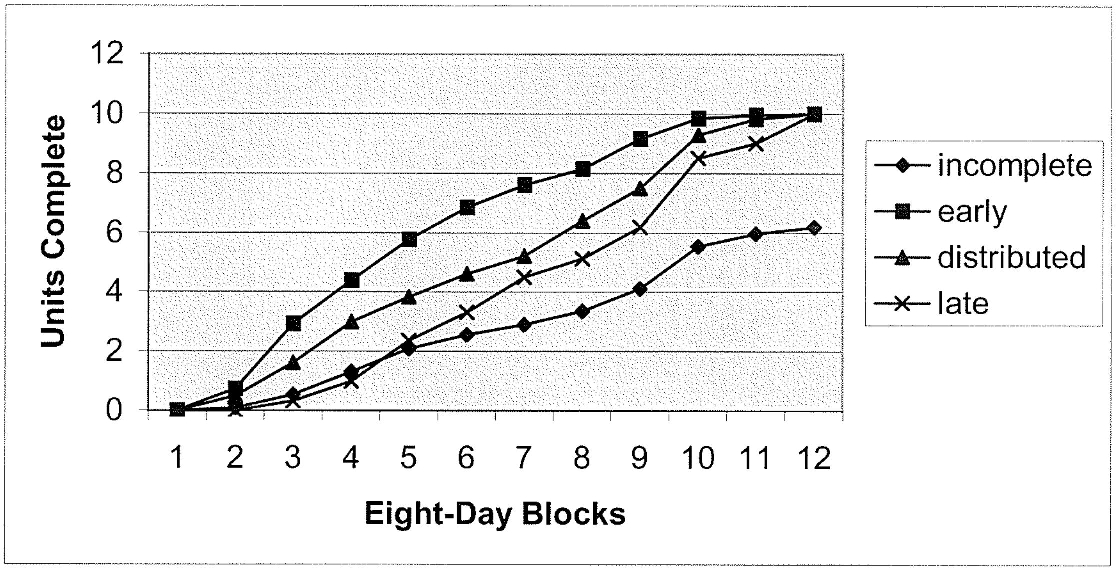


Figure 4. Mean progress through the units by each comparison group

Table 7 presents the overall analysis of variance (ANOVA) results for final exam scores in the Early-Massed, Late-Massed, and Distributed Progress, and Incomplete Groups. Note that overall differences were significant at the p = .002 level.

Table 7. ANOVA for the final exam scores of the Early-Massed, Late-Massed, and Distributed-Progress Groups, and Incomplete Group.

	Sum of Squares	Df	Mean Square	F	Significance
Between Groups	13.089	3	4.363	5.358	.002
Within Groups	74.911	92	.814		
Total	88.000	95			

Table 8 shows the result of the Dunnett T3 comparison of the Early-Massed, Late-Massed, and Distributed-Progress Groups with the Incomplete Group on the final exam. Note that the means of the groups that completed all the units were all higher than the mean of the Incomplete Group (as also seen in Figure 1). However, only the difference between the Early-Massed Group and the Incomplete Group was significant (p = .01).

Table 8. Dunnett T3 analysis of final exam scores of the Early-Massed, Late-Massed, and Distributed Progress Groups compared to the Incomplete Group.

Comparison Group	Progress Style	Mean Difference	Standard Error	Significance
Incomplete	Early massed	-0.9266221	.23218057	.001
	Late massed	-0.4420393	.22275162	.271
	Distributed	-0.5361236	.30023793	.390

Tables 9 and 10 present the overall analysis of variance (ANOVA) results for the scores on Midterms 1 and 2, respectively, in the Early-Massed, Late-Massed, Distributed Progress, and Incomplete Groups. Note that overall differences were not significant at the p = .05 level for Midterm 1, but they were for Midterm 2.

Table 9. ANOVA for midterm 1 scores of the Early-Massed, Late-Massed, Distributed-Progress, and Incomplete Groups.

	Sum of Squares	Df	Mean Square	F	Significance
Between Groups	5.751	3	1.917	2.144	.100
Within Groups	82.249	92	.894		
Total	88.000	95			

Table 10 – ANOVA for midterm 2 scores of the Early-Massed, Late-Massed, Distributed-Progress, and Incomplete Groups.

	Sum of Squares	Df	Mean Square	F	Significance
Between Groups	8.673	3	2.891	3.353	.022
Within Groups	79.327	92	.862		
Total	88.000	95			

Table 11 shows the result of the Dunnett T3 comparison of the Early-Massed, Late-Massed, and Distributed-Progress Groups with the Incomplete Group on midterm 2. Note that there was little difference between the performance on midterm 2 of the Late-Massed and Incomplete Group on midterm 2. Similarly, there was little difference between the performance on midterm 2 of the Distributed and Incomplete Groups on midterm 2. There was a larger, but still non-significant, difference between the Early-Massed Group and the Incomplete Group.

Table 11. Dunnett T3 comparison of midterm 2 scores of the Early-Massed, Late-Massed, and Distributed Progress Groups with the Incomplete Group.

Comparison Group	Progress Style	Mean Difference	Standard Error	Significance
Incomplete	Early Massed	-.4508197	.23472695	.303
	Late Massed	.3939600	.33990654	.831
	Distributed	-.3395080	.24250055	.655

Table 12 shows all exam scores for all groups. Note that the Early-Massed Group was consistently higher than all other groups on all three exams, and that the Incomplete Group was consistently lower than the mean of the other groups on all three exams.

Table 12. Final exam and midterm exams mean *z* scores (based on all four groups).

Progress Style	Final Exam Mean	Midterm 1 Mean	Midterm 2 Mean
Late massed	.0122782	.0632675	-.5364335
Distributed	.1063625	-.1210499	.1970345
Early massed	.4968610	.3293462	.3083462
Incomplete	-.4297611	-.2735197	-.1424735

The Pearson product moment correlation between number of peer-review points and final exam score was moderately positive ($r = 0.330$, $p = .001$). The correlation between progress score and peer review points was more positive ($r = 0.691$, $p = .000$).

The comparison of peer review points for each group showed two significant differences. These were between the incomplete group and the distributed group (p = .000) and the incomplete group and early massed progress group (p = .000). Table 13 shows the ANOVA table.

Table 13. Comparison of Each Progress Groups Peer Review Points.

Comparison Group	Progress Style	Mean Difference	Standard Error	Significance
Incomplete	Late Massed	-1.6782	1.54703	.853
	Distributed	-9.2484	1.66934	.000
	Early Massed	-11.9236	1.68033	.000

Discussion

The results of the Study 2 indicate that completing the units appears to be important, although there were students who did not complete all units and still did well on the final exam. Students who early massed had higher exam scores, and less variability in final exam scores. However, again since this was applied research the early massed progress may resemble more of an evenly distributed progress through the semester. Nevertheless, there were students in the Late-Massed and Distributed Groups who did as well as many students in the Early-Massed Group.

The correlations were as hypothesized: students who early massed earned more peer-review points and did better on the final exam. They appeared to have more time to peer review, as well as having completed the units before other students. Nevertheless, it is possible that motivation to do well on the final exam was somewhat lessened by the accumulation of peer-review points because there was a higher correlation between progress score and peer-review points than there was between peer-review points and final exam score.

In reviewing Figure 4 several behaviour trends of the students can be seen and speculated on. In the first 36 days there is exponential growth for all four progress groups. This may be sparked by excitement for the new course and semester. This may also be compounded by the need to complete units early in order to peer review and the lower workload due to no peer reviewing requirements early in the course (Kinsner & Pear, 1990). After the exponential growth the early massed group has a slightly logarithmic curve. This group completed many units early and their rate of unit completion slowed at about the midpoint of the course. At this time these students may have been spending more time collecting peer review points, particularly since this group showed to have higher peer review points than the late massed progress group and the incomplete group. The distributed group appears to steady pace through the course with a slight exponential growth near the end of the course. The late massed group also has an exponential growth near the end of the course. This may represent a push to complete units before the end of the semester as well as a push to complete units for the second midterm. The incomplete group appears to have two periods of exponential growth, the first previously described at the beginning of the semester and ending around the time of the first midterm. The second exponential growth occurs around the time of the second midterm, after that there are few units completed. The midterms during the semester may have contributed to these students behaviour pattern in completing units. These behaviour trends may represent typical student behaviour, with conscientious students and procrastinating students. The behaviour may also represents some effects of the CAPSI system. For example the slower completion of units later on by the early massed group may be a product of getting peer review points later in the semester.

General Discussion

Students who had higher peer reviewer points scored higher on the final exam. This may be because these students spent more time with the unit materials. This is partly supported by the high positive correlation with progress score and peer reviewer points. This appears to support (although it does not prove) the view that peer reviewing aids in the learning process, and that it is therefore a pedagogically useful course component.

The results of both studies seem to indicate that students can choose any method of progress without serious negative consequences to their exam performance and hence overall learning in the course; however, early-massed progress seems to be the most effective strategy and failing to complete all study units is the least effective. One should be cautious in inferring causal relations here. It could be, for example, that students who complete all units early are more motivated or conscientious overall, and those who fail to complete the study questions are less conscientious. Thus, there may be the other factor of motivation; viz., a tendency to be motivated or behave "conscientiously," which tends to result in a student completing units rapidly, earning many peer-review points (therefore reviewing the material more), and doing well on the final exam. As opposed to early mass progress and peer review points effecting final exam performance. Nevertheless, whether causally or diagnostically, it appears that completing all the units is a significant factor in performance on the exams. However, it also appears that students can work at their own pace—including procrastinating—without necessarily suffering a negative impact on their learning. Clearly, contrary to what was hypothesized, the Distributed Group was not consistently better than either massed progress groups.

A factor to keep in mind is the midterm exams. These exams might have acted to reduce student procrastination. Students may have tended to procrastinate only up to each midterm, thereby minimizing their procrastination and allowing them to more easily complete all the units. As we had seen in Figure 4, there is a slight trend toward this with both the late progress group and the incomplete units progress group. This trend of completing all units late is most pronounced for the final exam in comparison to the distributed progress group and the early massed progress group. This suggests that the midterm exams may have had an impact on individuals who choose to late mass progress.

The present study tends to supports the idea of allowing students to work at their own pace where this is administratively possible, such as in many online courses. It would still be recommended, however, for students not to procrastinate. Indeed, this study tends to support encouraging students begin working early in the course, so as to leave more time to interact with the material and study for the exams.

Conclusion

The evidence of the present research supports the hypothesis that allowing students to progress through the course at their own pace is not significantly detrimental to their exam performance. It did suggest that not completing all the units would hurt performance outcomes on the final exam. The positive correlations between peer review points and progress score and peer review points and final exam score seem to suggest that overall grade may be improved by early massed progress and getting more peer review points. Further research in this area examining other variables affecting exam performance is needed. Further research may also include the relationship of other variables such as GPA, peer review points, and midterm exam scores on final exam performance or final grade performance. As noted earlier, research on pacing in these courses is mixed; thus, continued research on pacing in these types of courses would benefit their structure and design in university programs.

References

Ainsworth, L. (1979). Self-paced instruction: An innovation that failed. *Teaching Psychology, 6*, 42-46.

Bangert-Drowns, R., Kulik, J. & Kulik, C. (1985). Effectiveness of computer-based education in secondary schools. *Journal of Computer-Based Instruction, 12,* 59-68.

Bahrick, H. & Phelps, E. (1987). Retention of spanish vocabulary over 8 years. *Journal of Experiential Psychology: Learning, Memory and Cognition, 13,* 344-349.

Brooke, R. & Ruthven, A. (1984). The effects of contingency contracting on student performance in a PSI class. *Teaching of Psychology, 11*, 87-89.

Brothen, T. & Wambach, C. (2001). Effective student use of computerized quizzes. *Teaching of Psychology, 28*, 292-294.

Canelos, J. & Ozbeki, M. (1983). Application of the Keller instructional strategy of personalized instruction for the improvement of problem solving learning in technical education. *Journal of Instructional Psychology, 10*, 61-69.

Crone-Todd, D., Pear, J. & Read, C. (2000). Operational definitions of higher order thinking objectives at the post-secondary level. *Academic Exchange Quarterly, 4(3)*, 99-106.

Dempster, F. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79*, 162-170.

Dempster, F. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist, 43*, 627-634.

Gray, P., Buerkel-Rothfus, N. & Yerby, J. (1986). A comparison between PSI-based and lecture recitation formats of instruction in the introductory speech communication course. *Communication Education, 35*, 111-125.

Hobbs, S. (1981). A comparison of student- and instructor-paced formats in the introductory psychology course. *Teaching of Psychology, 8*, 209-211.

Kinsner, W. & Pear, J.J. (1990). Chapter 7. A dynamic educational system for the virtual campus. Gattiker, U. & Larwood, L. (Eds). End-User Training. Walter de Grutyers Publisher.

Kinsner, W. & Pear, J. J. (1988) Computer-aided personalized system of instruction for the virtual classroom. *Canadian Journal of Educational Communication, 17*, 21-36.

Kulik, C., Kulik, J. & Bangert-Drowns, R. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 60*, 265-299.

Kulik, J., Kulik, C. & Bangert, R. (1984). Effects of practice on aptitude and achievement test scores. *American Educational Research Journal, 21,* 435-447.

Kulik, J., Kulik, C. & Cohen, P. (1979). A meta-analysis of outcome studies of Keller's personalized system of instruction. *American Psychologist, 34*, 307-318.

Keller, F. (1968). Goodbye, teacher..... *Journal of Applied Behavior Analysis, 1,* 79-89.

Lamwers, L. & Jazwinski, C. (1989). A comparison of three strategies to reduce student procrastination in personalized system of instruction. *Teaching of Psychology, 16*, 8-12.

Metalis, S. (1985). Effects of massed versus distributed practice on acquisition of video game skill. *Perceptual and Motor Skills, 61*, 457-458.

Murphy, H., McMichael, J. & Cariello, R. (1977). Correlating unit-quiz retakes and final examination performance. *Journal of Personalized Instruction, 2*, 49-51.

Nunnally, J., Duchnowski, A. & Knott, P. (1967). Association of neutral objects with rewards: Effects of massed versus distributed practice, delay of testing, age and sex. *Journal of Experiential Child Psychology, 5*, 152-163.

Pear, J. & Novak, M. (1996). Computer-aided personalized system of instruction: A program evaluation. *Teaching of Psychology, 23,*119-123.

Pickthorne, B. & Wheldall, K. (1982). A behavioral approach to teaching subsidiary physics to engineering students. *Educational Psychology, 2*, 79-85.

Rider, R. & Abdulahad, D. (1991). Effects of massed versus distributed practice on gross and fine motor proficiency of educable mentally handicapped adolescents. *Perceptual and Motor Skills, 73*, 219-224.

Ross, L. & McBean, D. (1995). A comparison of pacing contingencies in class using a personalized system of instruction. *Journal of Applied Behavior Analysis, 28*, 87-88.

Rothblum, E., Solomon, L & Murakami, J. (1986). Affective, cognitive and behavioral differences between high and low procrastinators. *Journal of Counseling Psychology, 33,* 387-394.

Schmitt, D. (1998). The introductory class in higher learning education: same old problems and new considerations. *The Behavior Analyst, 21*, 281-287.

Sherman, J. & Ruskin, R. (1978). *The Instructional Design Library, Volume 13.* Educational Technology Publications: Englewood Cliffs, New Jersey.

Solomon, L. & Rothblum, E. (1984). Academic procrastination: Frequency and cognitive-behavioral correlates. *Journal of Counseling Psychology, 31*, 503-509.

Steel, P., Brothen, T., & Wambach, C. (2001). Procrastination and personality, performance and mood. *Personality and Individual Differences, 30*, 95-106.

Tice, D., & Baumeister, R. (1997). Longitudinal study of procrastination, performance, stress and health: The costs and benefits of dawdling. *Psychological Science, 8*, 454-458.

Whitehurst, C. & Madigan, J. (1975). Slow learners in PSI courses: Do they learn less? *Journal of Higher Education, 46*, 55-62.

Webb, W. (1933). Massed versus distributed practice in pursuit-meter learning. Short Articles and Notes. Bartlett, F., Borovski, V., Kiesow, F., Matsumoto, M., Miles, W. & Wilcocks, R. (Eds.). *Journal of General Psychology, 8, 9*, 272-278.

Wesp, R. (1986). Reducing procrastination through required course involvement. *Teaching of Psychology, 13*, 128-130.

Wesp, R. & Ford J. (1982). Flexible pacing assists student progress in a personalized system of instruction. *Teaching of Psychology, 9*, 160-162.

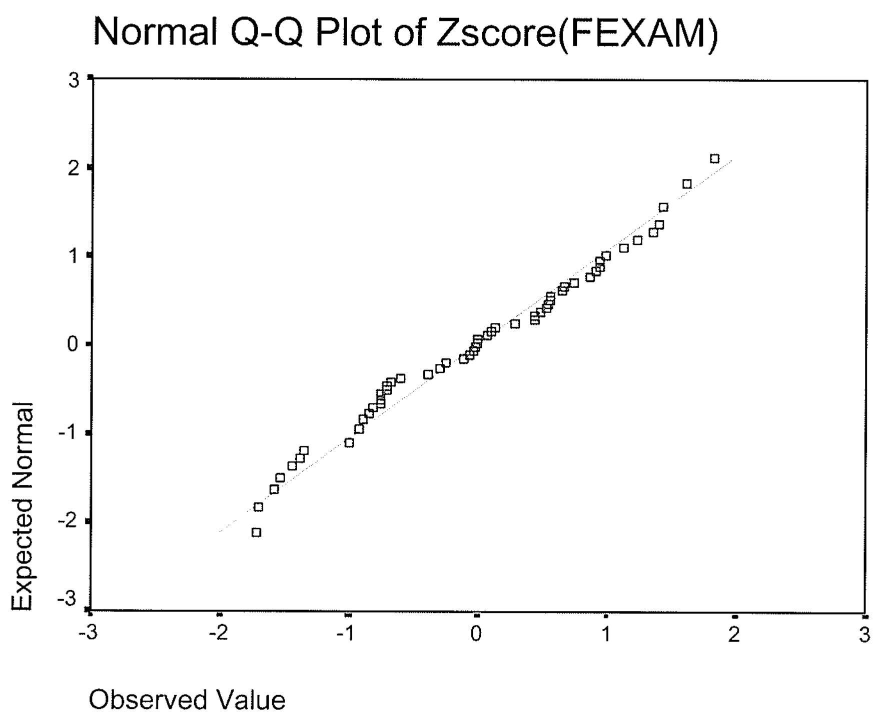
Zimmer, J. & Hocevar, D. (1994). Effects of massed versus distributed practice of test taking on achievement and test taking anxiety. *Psychological Reports, 74*, 915-919.

Appendix A

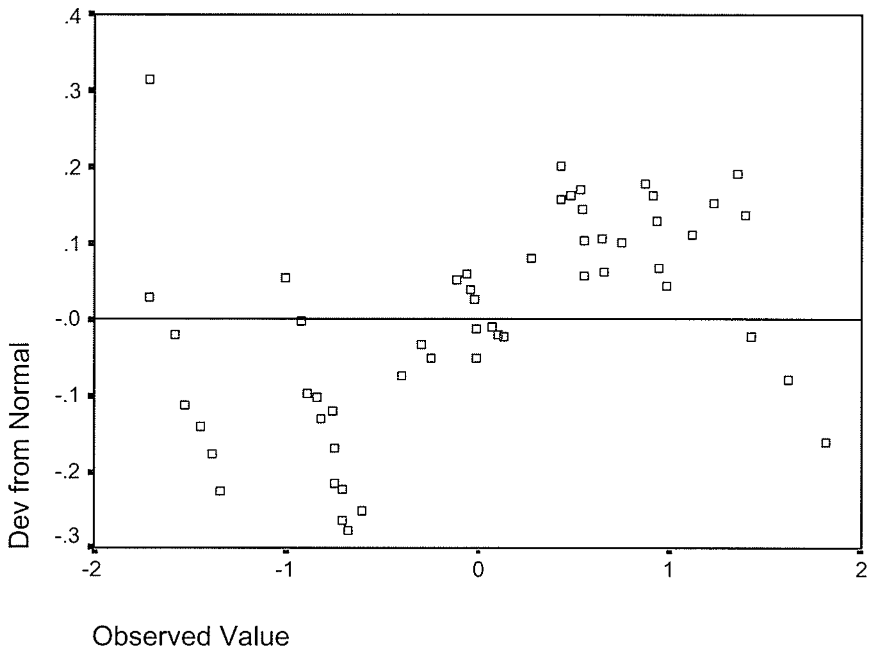
The Computer-Aided Personalized System of Instruction (CAPSI) combines personalized system of instruction with computer-mediated communications. The material in the CAPSI-taught course is divided into several units (e.g. 9). Typically the first unit covers the course procedures. Each unit has several study questions (e.g. 17 to 36 questions). Students progressed through the course by studying the written information and taking unit tests on the computer. Once the student is prepared to take a unit test he or she invokes the CAPSI program. The CAPSI program then emails the student a test, which contains a random sample of three questions from the study questions for that unit. When the student submits the test it is emailed to the instructor, teaching assistant, or two peer reviewers. A pass indicates that all three questions were answered thoroughly, while a restudy indicates that at least one of the questions was not adequately answered.

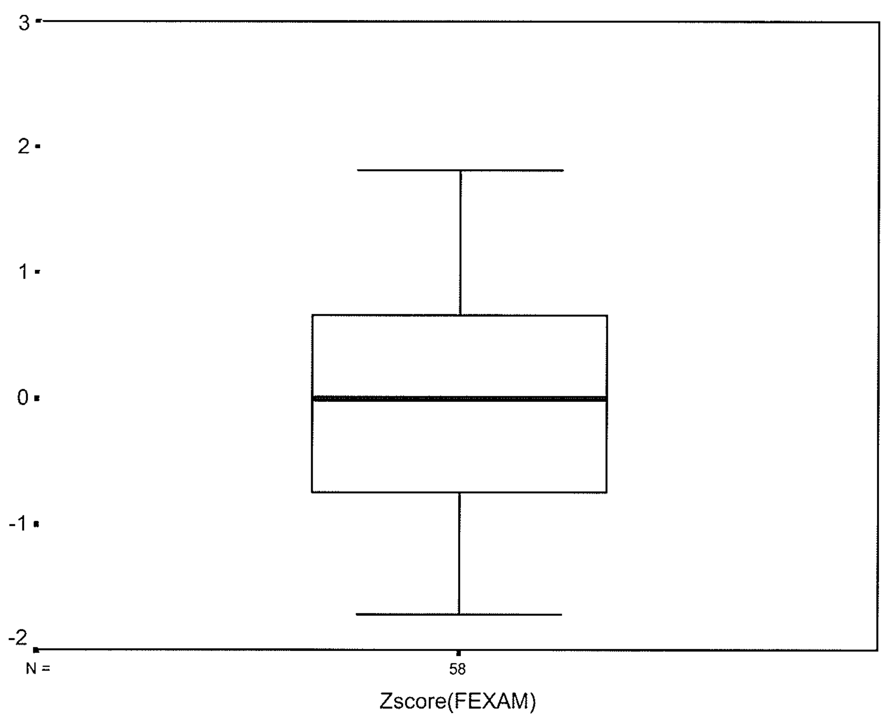
Appendix B

```
Zscore(FEXAM) Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     4.00       -1 .  5577
     4.00       -1 .  0334
    13.00       -0 .  6677777888999
    10.00       -0 .  0000012223
     7.00        0 .  0012444
    12.00        0 .  555566789999
     6.00        1 .  123344
     2.00        1 .  68

 Stem width:   1.00000
 Each leaf:       1 case(s)
```

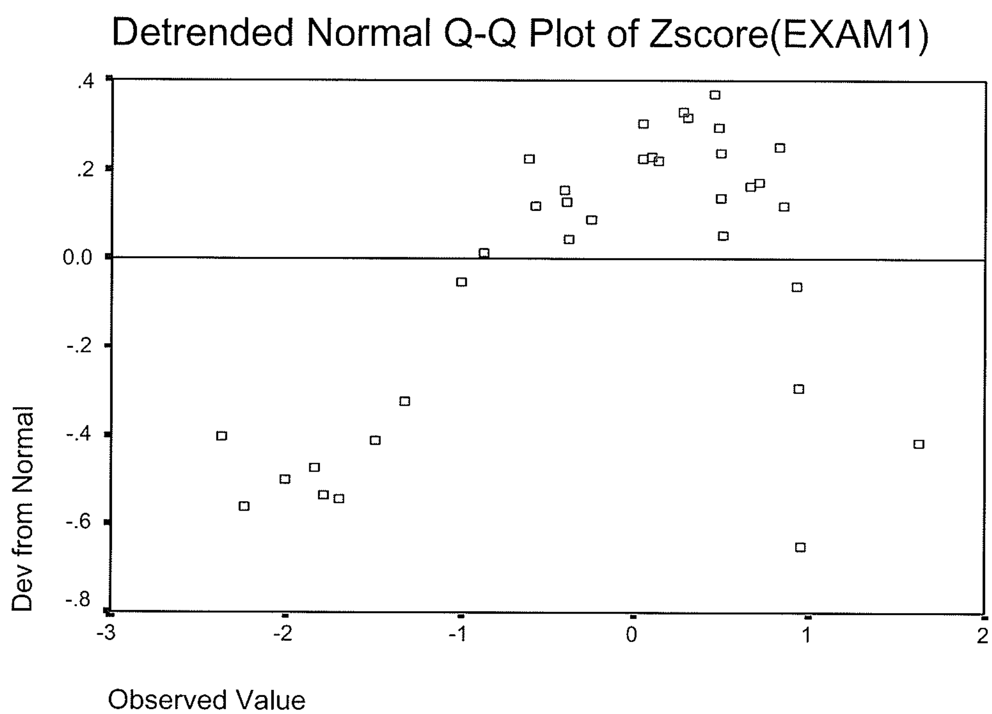


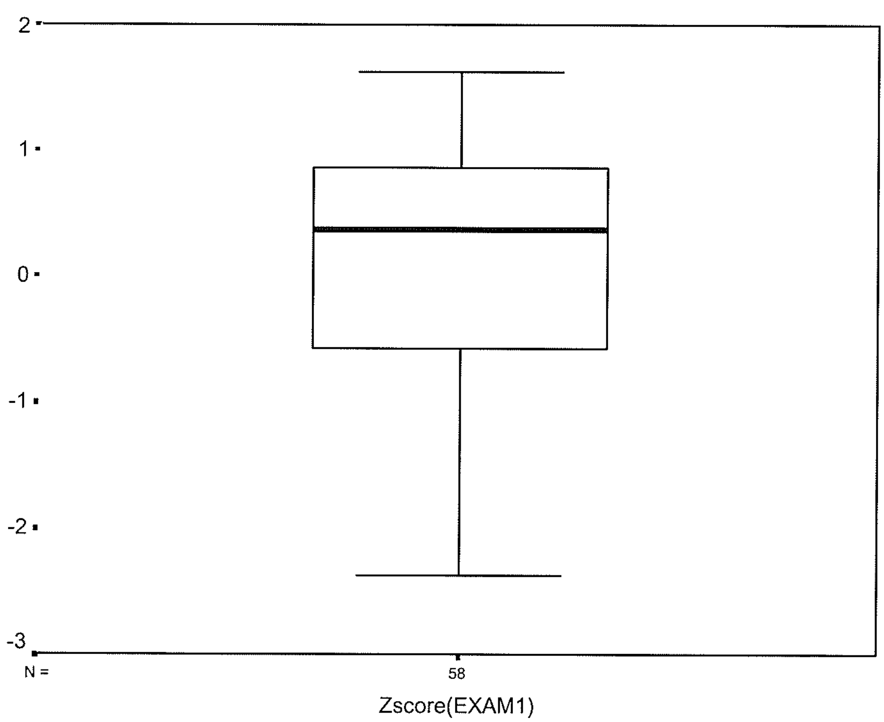
Detrended Normal Q-Q Plot of Zscore(FEXAM

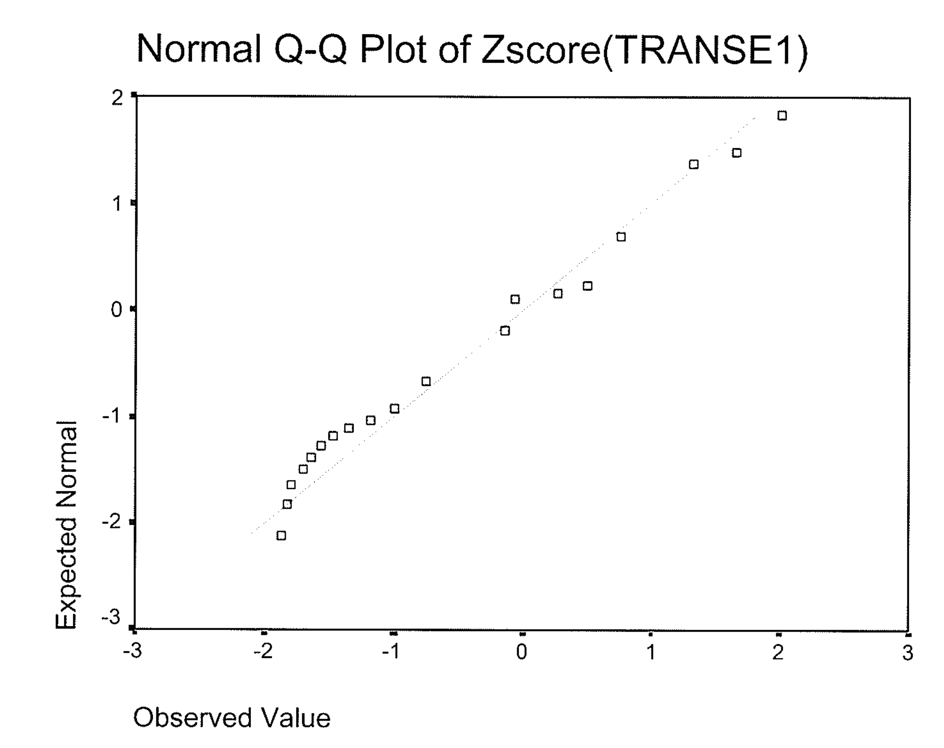


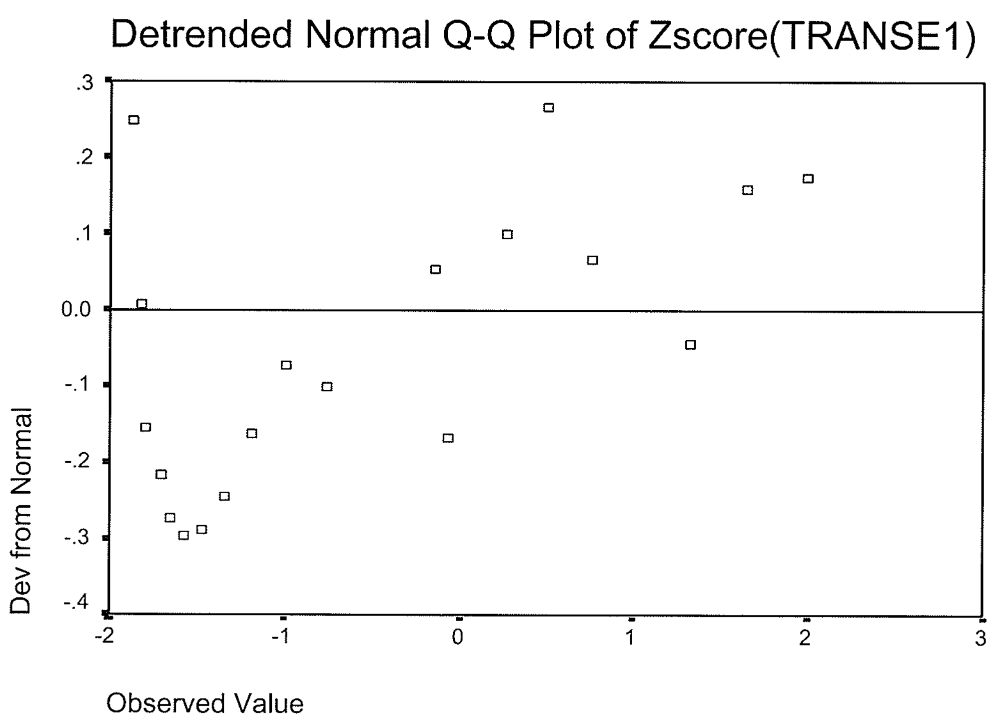


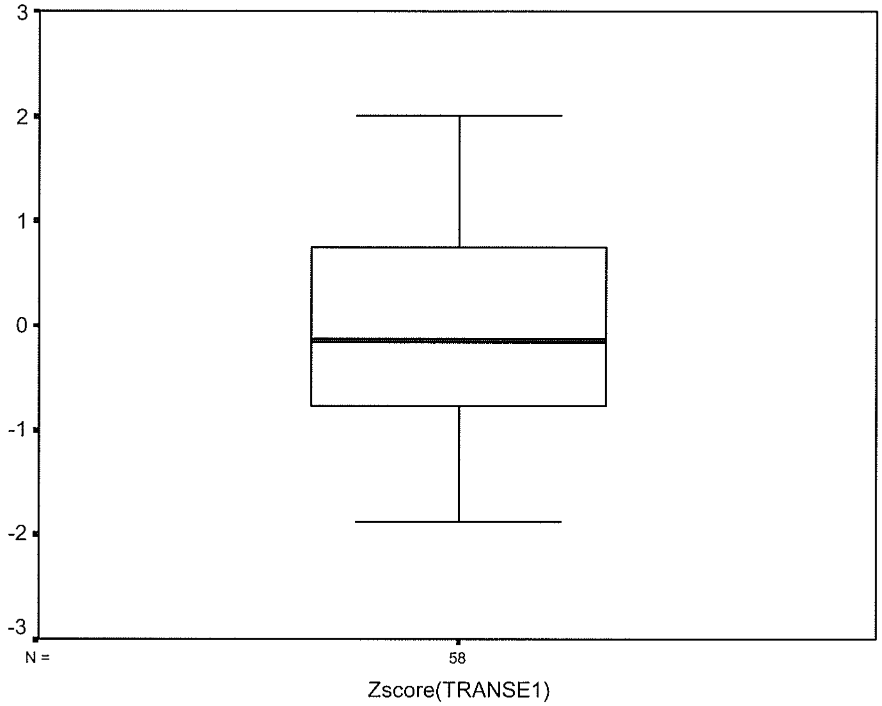
Appendix C


Detrended Normal Q-Q Plot of Zscore(EXAM1)
.4
.2
0.0
-.2
-.4
-.6
-.8
Dev from Normal
-3
-2
-1
0
1
2
Observed Value


2
1
0
-1
-2
-3
N =
58
Zscore(EXAM1)

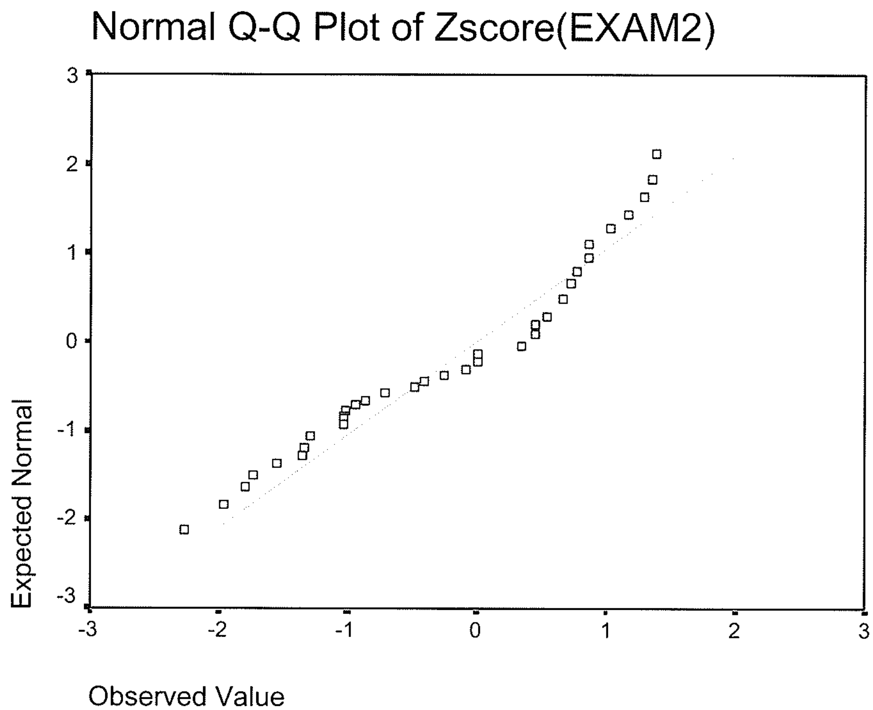

Normal Q-Q Plot of Zscore(TRANSE1)
Expected Normal
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
Observed Value

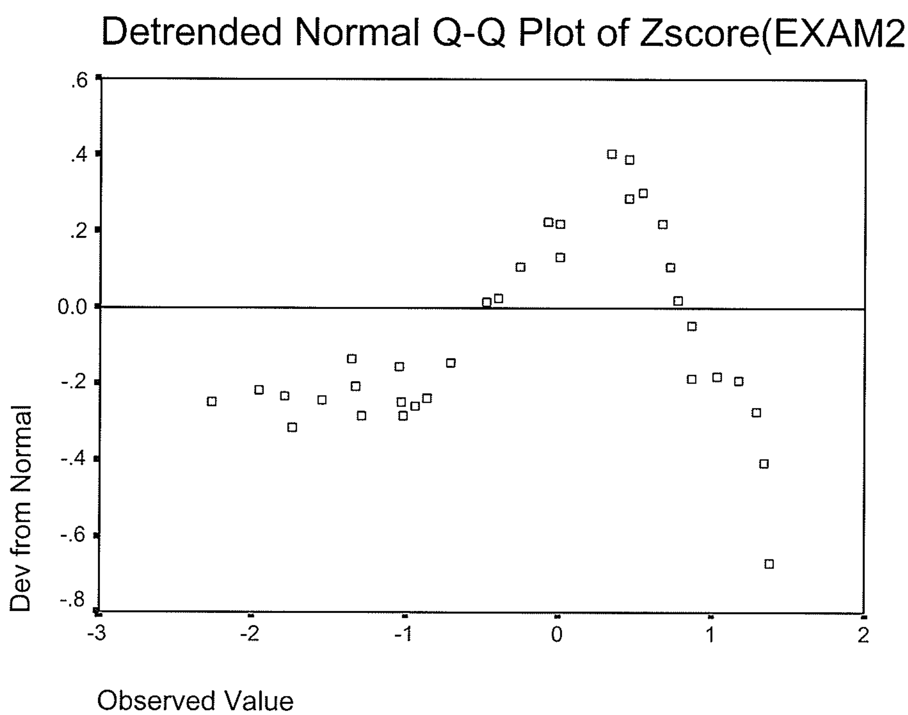

Detrended Normal Q-Q Plot of Zscore(TRANSE1)
Dev from Normal
.3
.2
.1
0.0
-.1
-.2
-.3
-.4
-2
-1
0
1
2
3
Observed Value

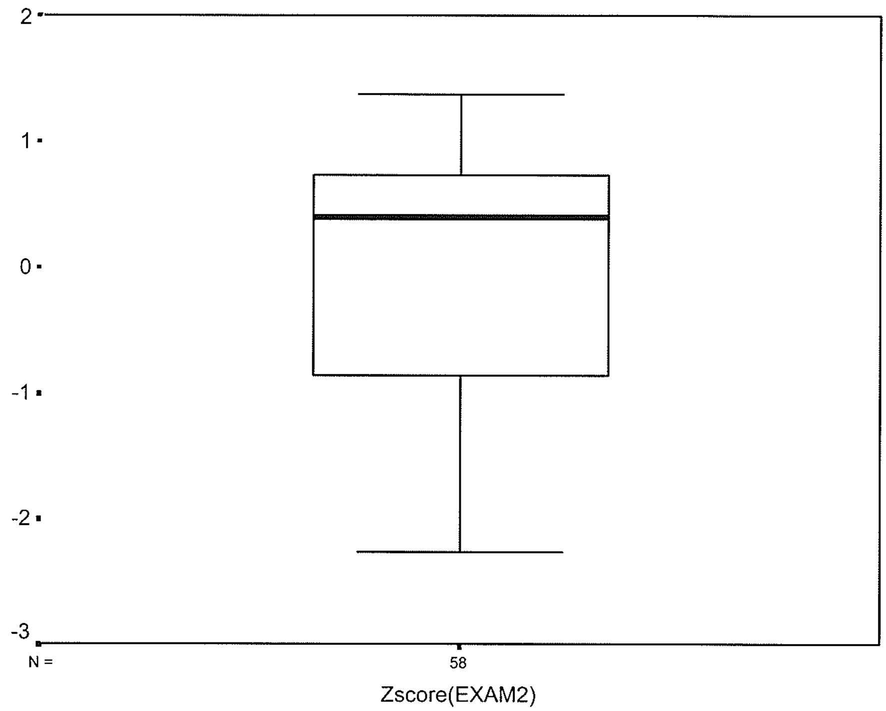


Graphs of Midterm Exam 2 Before Correction for Skewness

```
Zscore(EXAM2) Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     1.00       -2 .  2
     4.00       -1 .  5779
     8.00       -1 .  00002233
     4.00       -0 .  7789
     6.00       -0 .  002444
    11.00        0 .  00003344444
    18.00        0 .  555666666777778888
     6.00        1 .  011233

 Stem width:    1.00000
 Each leaf:        1 case(s)
```

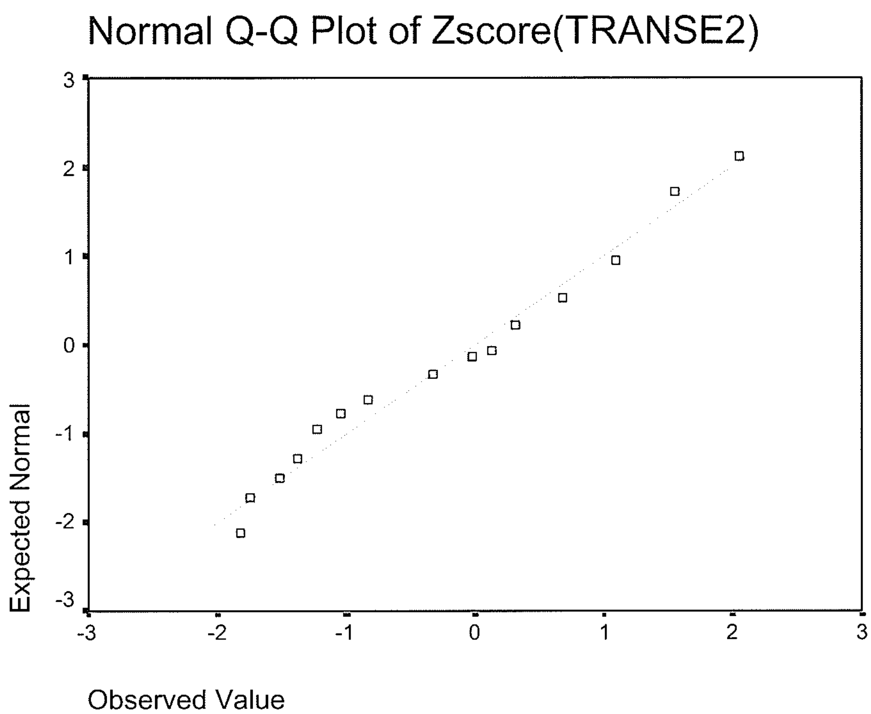

Normal Q-Q Plot of Zscore(EXAM2)
Expected Normal
Observed Value
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

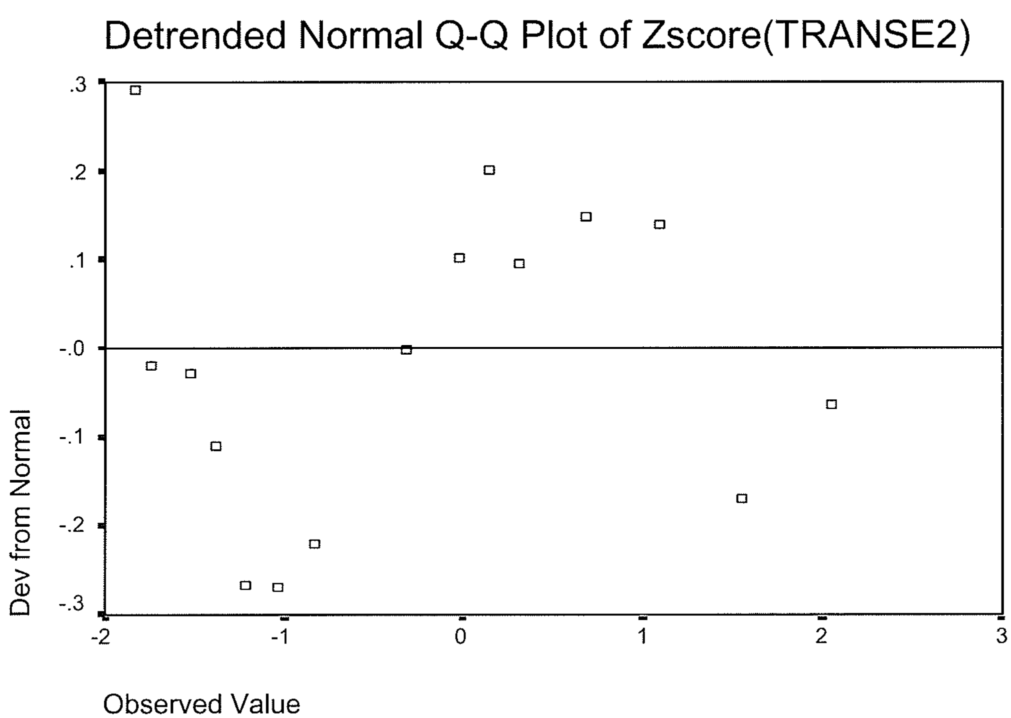

Detrended Normal Q-Q Plot of Zscore(EXAM2
Dev from Normal
Observed Value
.6
.4
.2
0.0
-.2
-.4
-.6
-.8
-3
-2
-1
0
1
2

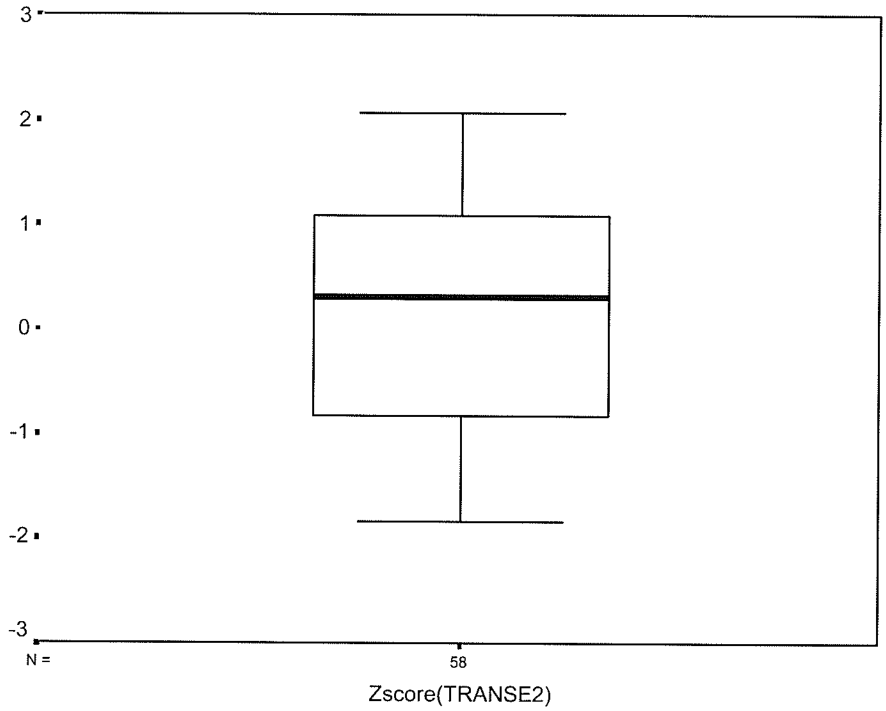


Graphs of Midterm Exam 2 After Correction for Skewness

```
Zscore(TRANSE2) Stem-and-Leaf Plot

 Frequency     Stem &  Leaf

     4.00        -1 .  5778
     9.00        -1 .  022222333
     5.00        -0 .  88888
     9.00        -0 .  003333333
    13.00         0 .  1333333333333
     2.00         0 .  66
    13.00         1 .  0000000000000
     2.00         1 .  55
     1.00         2 .  0

 Stem width:    1.00000
 Each leaf:        1 case(s)
```


Normal Q-Q Plot of Zscore(TRANSE2)
Expected Normal
Observed Value
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3


Detrended Normal Q-Q Plot of Zscore(TRANSE2)
Dev from Normal
Observed Value
.3
.2
.1
-.0
-.1
-.2
-.3
-2
-1
0
1
2
3


3
2
1
0
-1
-2
-3
N =
58
Zscore(TRANSE2)