Performance measures in courses using computer-aided personalized system of instruction

Crystal R. Springer, Joseph J. Pear *

Department of Psychology, Faculty of Arts, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

Received 29 December 2006; received in revised form 15 August 2007; accepted 3 September 2007

Abstract

Archived data from four courses taught with computer-aided personalized system of instruction (CAPSI) – an online, self-paced, instructional program – were used to explore the relationship between objectively rescored final exam grades, peer reviewing, and progress rate – i.e., the rate at which students completed unit tests. There was a strong positive correlation ($r = .68, p < .01$) between rate of progress and the amount of peer reviewing students did. This was predictable because peer reviewers had to be further along in the course than the students whose unit tests they reviewed. Students who completed all the units tended to obtain higher final exam scores than those who did not. For students who completed all the units there was little difference between the final exam performance of those who had a high progress rate and those who had a low progress rate. Considering all students together there was a moderate correlation between progress rate and final exam performance ($r = .36, p < .01$). In addition, there was a moderate positive correlation ($r = .33, p < .01$) between the amount of peer reviewing students did and their performance on the final exam. This correlation was substantially reduced ($r = .13, p < .20$) when rate of progress was partialed out. Thus, overall, students who progressed more rapidly through the course did more peer reviewing and learned more as measured by final exam performance. Interestingly, there were students who showed good learning without much participation in course related behaviors such as completing unit tests and peer reviewing. It is concluded that the CAPSI instructional program provides a good learning environment for students who utilize its components but it also accommodates other learning styles. In addition, there may be a distinction between students who complete all assigned units and those who do not regardless of their tendencies to procrastinate.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Computer-aided personalized system of instruction; Self-pacing; Procrastination; Early- vs. Late-Massed performance; Measure of educational performance; Essay tests; Student peer reviewers; Online instruction; Instructional technology

* Based on a thesis submitted by the first author in partial fulfillment of the requirements for the M.A. degree at the University of Manitoba. This research was supported in part by a Society for the Advancement of Behavior Analysis International Development Grant and a grant from the Technology Transfer Office of the University of Manitoba.

* Corresponding author.

E-mail address: pear@cc.umanitoba.ca (J.J. Pear).
1. Introduction

Meta-analyses of studies on personalized system of instruction (PSI), a teaching method developed by Keller (1968) and, therefore, also known as the Keller plan, have shown that it is more effective than conventional methods on a number of measures of learning (Kulik, Kulik, & Bangert-Drowns, 1990; Kulik, Kulik, & Cohen, 1979). The effectiveness of PSI may be attributed to the way that it incorporates established principles of learning (Sherman, 1982). Course material is divided into small sequential units; there are frequent unit tests based on clearly defined study objectives or study questions with rapid reinforcement of student mastery. Punishment is minimized. When a student does not demonstrate mastery on a unit test, the student writes another test on the unit after a sufficient amount of restudy time. Lectures are minimized or absent; instead, the text and other print material are the primary means of conveying course information and content to the students.

Because students can write unit tests at any time during a course, the method is self-paced. Traditionally, a PSI course relies on individuals called “proctors” to provide rapid reinforcement and feedback to students on their unit test performance (Sherman, Ruskin, & Semb, 1982). Proctors are students in a more advanced course who receive some course credit for their work with students in a less advanced course. The incorporation of proctors into the system has a dual benefit in that it provides (1) students in the lower level course with personalized attention in the form of frequent and rapid feedback, and (2) the proctors with the opportunity to strengthen their knowledge and gain fresh insights into the material.

Since 1984, the University of Manitoba has offered several undergraduate psychology courses that use a computerized version of PSI called Computer-aided personalized system of instruction (CAPSI; Kinsner & Pear, 1988, 1990; Pear & Crone-Todd, 1999; Pear & Kinsner, 1988; Pear & Martin, 2004; Pear & Novak, 1996). Rather than utilizing proctors – students external to the course – to provide feedback on unit tests, a CAPSI course uses peer reviewers – students within the course – who are selected by the computer to review or grade their classmates’ unit tests. Typically, unit tests – which consist of short answer and essay-type questions – are graded by two peer reviewers. Using two peer reviewers serves three purposes: (1) it provides additional feedback to the unit test writer; (2) it provides students with more peer reviewing opportunities; and (3) it provides an additional measure of quality control.

In addition to unit tests, CAPSI courses typically have midterm and final examinations. The examinations are distinct from the unit tests and are graded exclusively by the instructor. In addition, the final exam is weighted heavily because it is supervised whereas the unit tests and midterm exams are not. This is to guarantee that final grades are based on student learning.

The ability of computers to instantly access information regarding each student’s current level of mastery greatly facilitates the use of peer reviewers in a CAPSI course. To be eligible to be a peer reviewer on a unit, a student must have demonstrated mastery of that unit by having passed a test on the unit. Thus, the instructor or teaching assistant always grades at least the first few tests taken on each unit. When peers review a unit test both must assess it as a “pass” or it is considered a “restudy.” The criteria for a “pass” are that all answers must be correct by having no omission of information and no incorrect or inaccurate information (see Pear & Crone-Todd, 2002, for discussion of these criteria). Students can earn points toward their final grade by serving as peer reviewers on unit tests.

CAPSI allows the instructor to view all the unit tests written and all the comments and grades (“passes” and “restudies”) given by the peer reviewers. The instructor can then provide feedback or direction. CAPSI also contains a function for restricting any student from serving as a peer reviewer. In addition, there is an automatic process that sends appeals to the instructor.

In the present study, each CAPSI course consisted of ten study units, with approximately 15–25 questions per unit. Unit tests counted for 10% (1% each), the midterm exams together counted for 30%, and the final exam counted for 60% of the final grade. Each instance of peer reviewing added 1% to the final grade; thus, peer reviewing added bonus points to the final grade.

The present study focused on two major questions regarding the functioning of CAPSI:

1. Do students who procrastinate writing unit tests perform less well on the final exam?
2. Is peer reviewing correlated with how well students perform on the final exam?
The first question is of interest because of a common belief that procrastination – defined as delaying the completion of one’s work – is detrimental to learning. Substantiation of this belief would suggest either that self-pacing should be discouraged or that regular deadlines should be imposed. The literature regarding self-pacing is mixed. 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).

The second question is of interest because one of the justifications for assigning course credit to peer reviewing is that serving as a peer reviewer enhances learning. Evidence in favor of this view would suggest that peer reviewing is a useful course component.

2. Method

2.1. Subjects and data

Data from four undergraduate psychology courses (17.244, 17.245, 17.247 and 17.252) using CAPSI at the University of Manitoba during the fall semester of 2000 and 2001 were obtained from archived datasets. Course 17.244 (Behaviour Modification Principles) covered fundamental principles and procedures of behavior modification. Course 17.245 (Behaviour Modification Applications) covered designing, implementing, and evaluating applications of behavior modification. Course 17.247 (Learning Foundations of Psychology) covered basic conditioning principles and their relation to complex animal and human behavior. Course 17.252 (Orientations to Psychological Systems) compared and contrasted behavioral, cognitive, humanistic, and psychodynamic psychology with regard to fundamental issues in psychology.

The material for each course was divided into ten units, with 98 days of the course available to complete all units. Students proceeded through the units at their own pace, writing brief essay-type unit tests based on study questions. An online program, CAPSI, provided unit tests to students when they requested them, and submitted completed unit tests to the instructor, teaching assistant, or two peer reviewers who had passed the unit being graded. The priority in assigning a unit test to graders was to: (1) two students who had indicated their availability to peer review unit tests and who had earned the fewest number of peer review points; (2) the teaching assistant if two eligible peer reviewers were unavailable; (3) the instructor if two eligible peer reviewers and the teaching assistant were unavailable.

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 courses and were graded by the instructor. Students wrote the midterm exams online. The final exam was supervised in a computer lab at a scheduled time after the last day of classes.

The total number of students who wrote the final exam in all the courses was 97; of those, 61 had completed all 10 units in their respective courses. The numbers of students in the individual courses were: 17.244, N = 44; 17.245, N = 16; 17.247, N = 27; and 17.252, N = 10.

2.2. Method of analysis

The data analyzed in this study consisted of the progress score of the students, the amount of peer reviewing each student did, and each student’s objectively rescored final exam result.

A procedure for finding the area under a curve was used to obtain the progress score for each student. 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 of the courses (weekends and holidays included). This sum is called the progress score for each student. To take an impossible example for the purpose of illustration, if a student had completed all ten units in the first day of the course his/her progress score would be 10 \times 98 = 980. At the opposite extreme, if the student completed all 10 units on the last day, the progress score would be 0 \times 97 + 10 = 10. Thus, progress scores for this study fell between 10 and 980.
Progress scores were divided into the three comparison groups: Early-Massed, Late-Massed, and Distributed. Early-Massed progress was defined as scores above 507; Distributed progress as scores between 393 and 507; and Late-Massed as scores below 393.

To ensure objectivity in the final exam score as a dependent variable, performance on the final exam was rescored with clear operational definitions. Inter-observer reliability measures were taken on scoring the exams for each course. The level of agreement of the exam scores was evaluated with the Kappa statistic. This statistic corrects for chance agreement between scorers. The range of the Kappa scores was from 0.836 to 0.936, which is considered high.

3. Results

The final exam scores for each course 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.

Fig. 1 shows box plots of the distributions of final examination scores for the Late-Massed, Distributed, and Early-Massed groups, and for the students who did not complete all ten units. Note that the Early-Massed Group had the highest mean $z$ score, the Distributed Group had the next highest, the Late-Massed Group had the next highest, and the Incomplete Group had the lowest. Nevertheless, there was a considerable amount of overlap between the groups. Note also that the massed groups had less variability than the Distributed and Incomplete groups. It may be that the massed groups were more homogeneous than the other two groups.

Table 1 shows the means, in terms of $z$ scores, on the final examination of the four groups. Note that these means are consistent with the medians shown in Fig. 1: the Early-Massed Group had the highest mean, the Distributed Group the next highest, the Late-Massed the next highest, and the Incomplete Group had the lowest.

Table 2 shows the ANOVA on the final exam for the four groups. Note that the $F$ is significant well beyond the $p = .01$ level. Thus, there were clear differences in the mean performances on the final exam of at least two of the groups.

Table 3 shows the result of a Dunnett’s T3 statistic comparison of the Early-Massed, Late-Massed, and Distributed Groups with the Incomplete Group on the final exam. Dunnett’s T3 was applied because it is used for multiple comparisons where one group is compared to all other groups without comparing the other groups to one another, and it does not assume equal group size. Levene’s statistic was 3.802 ($p = .013$), which

![Boxplots of the final exam scores of each of the indicated groups.](image)
indicated unequal group variances; however, Dunnett’s T3 does not assume homogeneity of variance. Note that only the difference between the Early-Massed Group and the Incomplete Group was statistically significant (p = .01).

Table 4 shows the inter-correlations between the three main variables in this study: progress score, number of peer reviews, and final exam score. Note that all of these correlations were positive and statistically significant.

The moderately positive correlation of progress score with final exam score is consistent with the data in Fig. 1 and Tables 1–3. This correlation appears to be largely due to those students who did not complete all the units. When their data are omitted from the analysis, the correlation is slightly negative and non-significant (r = −.17, p > .05).

The moderately positive correlation between peer review points and final exam score (r = 0.331, p = .001) might reflect the correlation between peer reviewing and progress score, because progress score was positively correlated with peer reviewing points and with final exam score. The correlation between peer review points and final exam score was changed when progress score was partialed out (r = .129, p = .209).

4. Discussion

The results indicate that students who begin their work on unit tests early obtain higher final exam scores than students who do not complete all the units. Within those that completed all the units, there are small, statistically insignificant differences and considerable overlap between the groups. This supports 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 work at their own pace does not deleteriously affect the exam results of many students – provided that they complete all the units. There were students in the Late-Massed, Distributed, and Incomplete Groups who did as well as many students in the Early-Massed Group. If there had not been an arbitrary cut-off date for completing all the units, the differences among the groups may have been even smaller.

The correlations indicated that in addition to tending to do well on the final exam, students who massed early earned more peer review points. It appears that even though CAPSI assigns unit tests for review to students who had fewest peer review points, completing units early – and thus being ahead of other students – provides a distinct advantage in being assigned to peer review and thus to earn peer review points.

There are several possible reasons that students who progressed most rapidly through the units tended to perform better on the final exam than those who progressed less rapidly. The first students who completed a unit test had their tests graded by the instructor or teaching assistant, and it might be argued that these students had the more expert feedback. More motivated students might also prefer to have their unit tests graded by an expert, and thus would possibly race to be the first to finish. Moreover, students who were last to finish would have fewer opportunities to peer review, which could detrimentally affect their performance. Future studies will be needed to tease out the impacts of these variables. It might be expected that students going into the final exam with more peer review points would be less motivated to do well on the final. Despite this supposition, students who earned more peer review points tended to do better on the final exam. 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. Thus, scores on the final exam may have underestimated the learning of students who peer reviewed a lot.

Nevertheless, students who had higher peer reviewer points scored higher on the final exam. However, this became statistically insignificant when progress scores were factored out through partial correlation. Thus, it cannot be concluded from these data that peer reviewing per se resulted in improved scores on the final exam. It may be, for example, that the more motivated students tended to start work on the units earlier and to do better on the final exam. The fact these students tended to do more peer reviewing may have been the result of starting working on the units earlier, because the program does not allow students to review tests on a unit until they have passed a test on that unit. It is also possible that the points for peer reviewing aided in the motivation to complete the units early, thereby moving students more quickly through the material and resulting in higher final exam performance.

A factor to keep in mind is that the midterm and final exams may have tended to limit the self-pacing aspect of the course. For example, students may have tended to procrastinate only up to each midterm exam, thereby minimizing their procrastination and allowing them to more easily complete all the units. However, even with the midterm exams there were still students who completed a minimal number of units, indicating that they were procrastinating throughout the entire semester. No specified number of units was required to be completed before each exam. Thus, the exams clearly did not completely restrict self-pacing.

It should also be noted that the exams might have caused the group distributions to be somewhat arbitrary. For example, students who massed early may still have distributed their review of the material when preparing for the exams.

Overall, however, this study tends to support allowing students to work at their own pace where this is administratively feasible. Nevertheless, it would still be recommended that students not procrastinate. Indeed, this study tends to support encouraging students to begin working early in the course, so as to leave more time to interact with the material and to study for the exams. In addition, there may be a distinction between students who complete all assigned units and those who do not – regardless of their tendencies to procrastinate. That is, there may be two distinct categories of procrastination.

5. Conclusions

The evidence of the present research suggests that allowing students to progress at their own pace is not detrimental to their final performance. The results also suggest, however, that not completing all the study
units can hurt performance on the final exam. The positive correlation between peer review and progress score and between peer review and final exam score suggest beginning work early and doing more peer reviewing improves the overall performance. Further research in this area examining other variables affecting exam performance is needed. Research that directly manipulates the pace at which students progress through a course by imposing deadlines, and that directly manipulates the amount of peer reviewing students do or the amount of credit earned for reviewing, should be particularly informative as to whether the effects found in this study are causative or only correlational. Further research might also look at the relationship of other variables, such as overall grade-point average, that might be correlated with performance on the variables examined in this study. As noted, earlier research on self-pacing is mixed; thus, research is needed on the conditions under which self-pacing is beneficial, detrimental, or benign.

References