JLD Journal of Learning Design
DESIGNING FOR EFFECTIVE LEARNING

AN INVESTIGATION OF COMPUTER GENERATED KNOWLEDGE RETENTION ACTIVITIES IN COMPUTER-BASED TRAINING WITH ADULT LEARNERS

Doug Reid
Faculty of Education
University of Wollongong, AUSTRALIA
Doug_reid@uow.edu.au

D. Michele Jacobsen
Faculty of Education
University of Calgary, CANADA
dmjacob@ucalgary.ca

Larry Katz
Faculty of Kinesiology
University of Calgary, CANADA
katz@ucalgary.ca

Abstract

The goal of this investigation was to evaluate the impact of training and the effectiveness of different types of knowledge retention activities delivered by computer-based training programs. This study focused on a computer-based learning system called the Profound Learning Delivery System (PLS). PLS is an application designed to improve the content knowledge retention of adult learners who are completing computer-based training.

This study used a pretest-posttest experimental design to compare adult learners' knowledge of Microsoft Outlook ("Outlook," 1997) before and after a computer-based training session. Participants were trained using two different computer-based instructional programs; a commercially available software program matched for comparison purposes and PLS. This comparison involved three different formats for post-instruction retention activities that were; no review activities, user generated review activities, and program generated retention activities. Results indicate, there was a significant difference between the groups 60 days after training. This result demonstrated that PLS has potential worth exploring.

Keywords
Knowledge retention, adult learning, computer-based training

Introduction

Computers and the Internet are revolutionizing the way people communicate and learn (Jackson, 2004). A better understanding of how adults retain knowledge from computer-based training is vital in an increasingly competitive business where it is essential to develop time and cost-effective methods of training employees in order to enhance their performance.

There is a paucity of research literature regarding knowledge retention and computer-based training—specifically instructional design and how it increases knowledge retention in technology-based instruction (Caple, 1996; Fletcher-Flinn, 1999; Streatmans & Eggen, 1989). Determining the long-term effects of instructionally designed and learner designed knowledge retention activities may provide useful information for the design of computer-based instruction programs in the future. A review of the

literature shows that many studies are knowledge retention or computer-based instruction studies; few are both. There continues to be a definite need to explore these issues along with adult learning in the workplace as the trend toward more computer-based training continues (Caple, 1996).

Purpose of study

This study evaluated the effectiveness of different knowledge retention activities performed during and after computer-based training (CBT) sessions. This study compared content knowledge retention of adults trained on two different instructional programs. The investigation was designed to query which type of retention activity was most effective for adult learners in the workplace as measured directly after training as well as 30 and 60 days after the initial computer-based training.

Literature review

A great deal of the literature on memory and recall focuses on how people process information (Okolo & Ferretti, 1996; Smith, 1998; Son, 2004). It has been argued that recall is mainly influenced by how new information is integrated with material already stored in long-term memory (Son, 2004; Sprenger, 1999). It has also been argued that learner behaviour is the most important factor during and following initial contact with new information (Theide & Dunlosky, 1999).

An aspect of learner behaviour is student self-discipline. A model of self-regulated learning was set forth (Theide & Dunlosky, 1999) that had three components: planning, discrepancy reduction, and working memory constraints. Participants regulate their learning by setting a desired goal for learning an item. They monitor how well they feel their learning is progressing and adjust their behaviour with the ultimate goal of learning the material in mind. An offshoot of self-regulated study is daily repetition. Daily repetition of important information is another strategy for building long-term memory (Carrier & Pashler, 1992; Lieberman & Linn, 1991; Sinclair, Healy, & Bourne, 1997; Sprenger, 1999).

Feedback has been found to be very important for learners during their instructional session (Baylor & Chang, 2002; Clariana, Ross, & Morrison, 1991; Collis, Boer, & Slotman, 2001; Kneebone, Scott, Darzi, & Horrocks, 2004). There are significant benefits to the user for even a minimal amount of feedback over no feedback at all (Clariana et al., 1991; Mathan, 2004; Smyth, 2004). Feedback in computer-based training has many different forms including timing, purpose and adaptiveness.

The timing of feedback is featured throughout the literature (Baylor & Chang, 2002; Kulik & Kulik, 1988). The variable of timing concerns when the user receives feedback during instruction, after instruction, during evaluation, and after evaluation (Baylor & Chang, 2002). Another aspect of feedback timing includes the possibility of time-delayed feedback designed to allow the user an opportunity to think about the question that triggered the feedback. It has been argued that feedback immediately after user response was best for most instructional situations (Kulik & Kulik, 1988).

The purpose of feedback is important for the format in which it is presented to the user. Evaluative feedback can be as basic as a correct or incorrect message, or it may include quantitative data such as the number of correct versus incorrect responses or the time it took to complete the training. Instructional feedback has a different purpose, therefore explanations and greater detailed information may be provided to the user (Boston, 2002). This type of feedback might lead to further questions or data to allow the user to explore a topic of interest or review a topic of difficulty.

Adaptive feedback is one of the strengths of computer-based training (Bjorner, Kosinski, & Ware, 2003; Embretson, 1996). An examination of computerized adaptive testing and Item Response Theory has been carried out by a number of researchers (Streatmans & Eggen, 1989; Ware, Bjorner, & Kosinski, 2000). Item Response Theory allows a computer application to have knowledge benchmarks. The program skips questions when the learner takes the test. These benchmarks are based on the assumption that if the learner can answer a question correctly, then the learner can answer all the previously skipped questions correctly (Bjorner et al., 2003). When the learner answers a

question incorrectly, the program skips back in the question list to allow the learner to answer a previously unanswered question and to reduce the skip interval.

Some of the main benefits of adaptive testing include individualization, difficulty level, test length, and question security (Embretson, 1996). An adaptive test adapts itself to the ability of each person taking the test. Therefore, each test has an individualized difficulty level rather than a generic difficulty targeted at the average ability level of people in the test group. Adaptive testing allows a person to answer fewer test questions, thus potentially allowing the test to be completed in less time. It also helps improve the security of the test because each person takes a potentially unique test. Other benefits of computer-based adaptive testing include on demand test delivery and computer-based test marking.

Method

This study employed a pretest-posttest experimental design with a convenience sample of adults in a corporate environment. Adult participants were randomly assigned into one of three groups, and provided with computer-based training for portions of Microsoft Outlook. The formats of evaluated knowledge retention activities were quite different for the 60 days following the initial training. The three groups differed as follows:

1. Focus Q with no review activities
2. Focus Q with user generated review activities
3. Profound with program generated retention activities

This material will be presented through a more detailed explanation of the participants, the software used, participant training, review and the testing process.

Participants

All participants were adult employees of a multi-national company that had an office in a city in a western country. Participants volunteered for the study and were given no incentive to participate beyond access to additional training on the software.

The participants completed a demographic survey that showed that they were all regular computer users. Generally, participants had several years experience with computers and they all used computers every day, at home, in their workplace or both. Ninety seven percent of the participants had home computers, 91% had Internet connections. Of the 42 participants who completed the initial training, 32 (20 males and 12 females) finished the study. The mean age of the participants was 42 years, with a range from 31 to 59 years old. The participants responded that on average they spend 6.7 hours a day using computers. For experience using computers, the range was from five years to 38 years of computer experience.

Software

There were two different types of software used for CBT purposes in this study, PLS and Focus Q. The Profound Learning Delivery System (PLS), designed by Profound Learning Systems Inc., is an Internet-based instructional software program designed to individualize content retention activities after an instructional session ends ("Profound Learning Delivery System," 2005). The knowledge retention aspect of the program is run by the PLEngine, which modifies the retention questions to suit the individual learner, and provides feedback to the user about their achievement. Information retention was the focus for the PLS in its training and retention activities. While the daily retention sessions were scheduled to be only five minutes long, they were individually adaptive to the learner's performance. Focus Q, available in both CD-ROM and web based formats, was an instructional software program designed with an adaptive learning capability. During instruction, Focus Q used an adaptive testing algorithm, which shortens testing time while determining the learner's mastery or non-mastery of the course.

There were a number of similarities between the PLS and the Focus Q programs used in this study. Both programs had text, audio and graphical aspects to their presentation of information. Both programs included a modular adaptive component in which designers can insert the instructional content to be learned. Focus Q and PLS both included built-in testing components and allowed users to control the pace in which instructional content is presented.

There were a number of differences between the PLS and the Focus Q programs used in this study. The main difference was the individualized content retention activities which were in the PLS. The two programs presented the same content material with slight variations in much the same way that two teachers would teach the same content in their own style. The PLS presented the content with an aim to reuse portions of it in the retention activities, while Focus Q did not.

Training

Training consisted of having the participants attend a full day session at an off-site computer lab. The participants volunteered to learn how to use the computer application Microsoft Outlook. Training was done over three days, with one third of the participants undergoing training each day. To ensure that each participant could complete the entire training session in a single day, only certain functions of the application were included in the computer-based training programs. These functions included: notes, tasks, address book, journal, contacts and custom views. Since the training and the posttest were self-paced, the time taken to complete the training varied from 4 to 7 hours depending on the pace of individual participants.

The participants were randomly split into three groups: Focus Q without Review, Focus Q with Review, and Profound. Groups Focus Q without Review and Focus Q with Review used Focus Q while the Profound group employed PLS. The two Focus Q groups worked through the same training. The difference between the groups involved retention activities during the post-training portion of the study. Both Focus Q groups were given a CD-ROM with the Focus Q program on it and instructions on how to install it onto their computers at home and at their work place. The Profound group logged into and used PLS during training.

Review

The three groups had different directions for what they were expected to do during the 60 days following the training. Focus Q with Review and Focus Q without Review were given a different set of instructions following the training. The Focus Q with Review group was instructed to independently review the CD-ROM version of the Microsoft Outlook training program for approximately 5 minutes a day for 60 days while Focus Q without Review group was not required to do any review activities. The Profound group was required to log in to the PLS for the 5 minutes of retention activities generated by the program each day.

Testing

Participants completed a content evaluation test on four separate occasions. The pretest was carried out before the participants received any training. The second test was done on their training day directly after the training session. The third content evaluation test occurred approximately 30 days after the initial training session and the final evaluation came approximately 60 days after the initial training session.

The content evaluation test consisted of 100 knowledge questions about the components of Microsoft Outlook they had been trained in. The 100 content questions included 77 true or false questions, 7 multiple choice questions with one correct answer and 16 multiple choice questions with possibly more than one right answer.

Results

Based upon test performance over time, this study attempted to determine which type of retention activity is most effective for adult learners in the workplace. The test items were the same for each administration, which allowed the test to act as a review constant.

There was a high degree of variability in variances across groups, as an examination of Table 1 will demonstrate. Moreover, the Profound group shows a marked decrease in variance over time that is not evident in the other two groups. Table 1 provides the means and standard deviations by group for the pretest posttest, and 30 day and 60 day retention tests.

Table 1: Means and Standard Deviations for pretest and posttest data for all participants who completed the study.

			Pretest	Posttest 1	30-day	60-day
Focus Q without Review	N = 11	Mean	64.27	77.64	74.64	70.09
		S.D.	5.18	6.58	4.76	5.77
Focus Q with Review	N = 12	Mean	66.42	79.17	75.58	72.67
		S.D.	6.79	7.59	7.17	8.27
Profound	N = 9	Mean	68.33	86.11	81.89	79.33
		S.D.	7.21	5.06	4.14	2.83
Total		Mean	66.22	80.59	77.03	73.66
		S.D.	6.41	7.33	6.30	7.16

Plots of individual performances by group overtime are provided in Figures 1,2, and 3. It would appear that the Profound group have a distinct retention pattern over time that is not evident in the Focus Q groups.

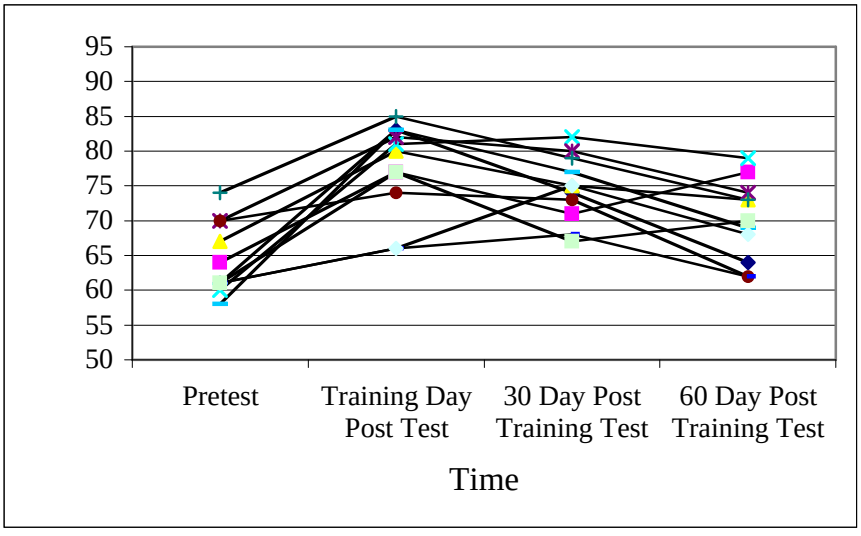

Figure 1 – Graphical representation of all participant test scores in the Focus Q without Review group.

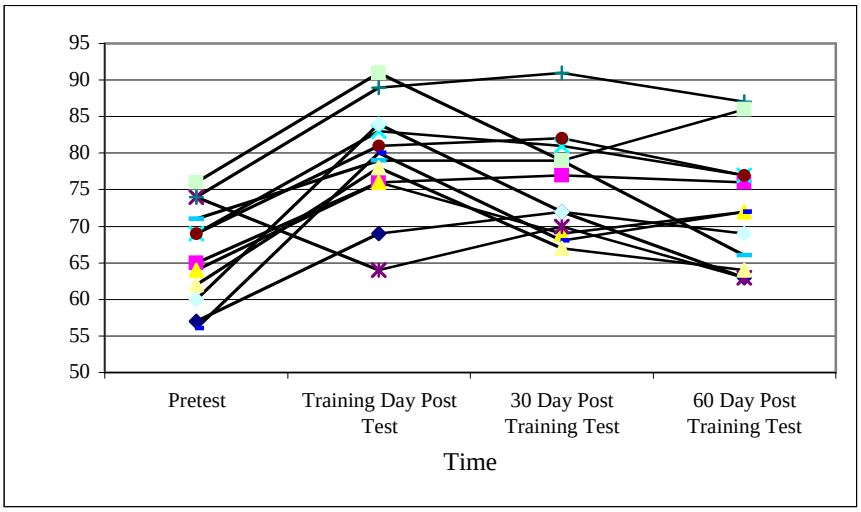

Figure 2 – Graphical representation of all participant test scores in the Focus Q with Review group.

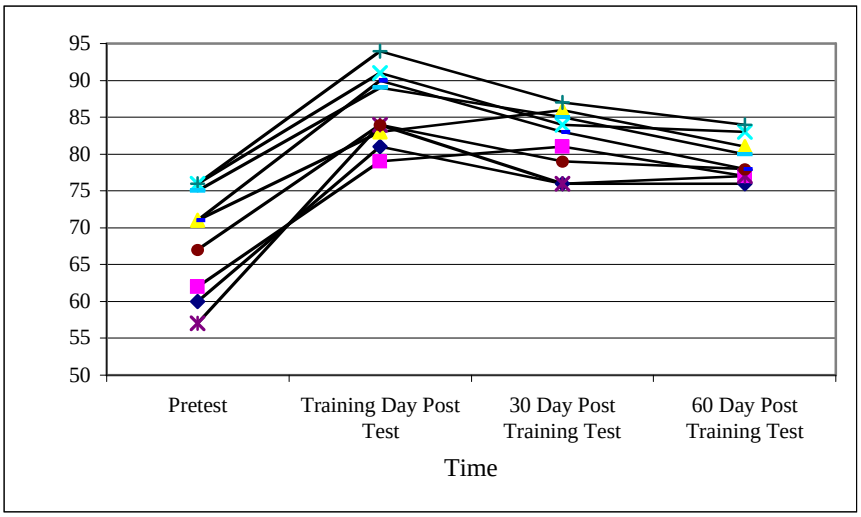

Figure 3 – Graphical representation of all participant test scores in Profound Group.

An Analysis of Variance for homogeneity of Variance shows a significant difference in variances between groups. When repeated measures of ANOVA were performed on the content evaluation results, there were significant differences between groups (See Table 2).

Table 2: A two way Analysis of Variance with repeated measures for pretest, posttest, 30 day and 60 day retention tests.

Effects	SS	Df	MS	F	P
Groups(Grp)	1748.47	2	874.24	10.35	.001
Residual	2364.28	28	84.44		
Time (T)	3683.14	3	1227.71	69.92	.001
Grp x Time	111.80	6	18.63	1.06	NS
T x Residual	1527.57	87	17.56		

One explanation for the non-significant group by time interaction may be the significantly high variability in the two Focus Q groups. In order to evaluate the impact of this variability, an individual analysis was performed for each testing period and the results are shown in Table 3. Table 3 provides

an analysis of the variance for group means for each of the four testing periods (Pretest, Posttest 1, 30-day post training and 60-day post training). The pretest results suggest that there were no significant differences in content knowledge among the three groups prior to training. There were statistically significant differences between groups after Posttest 1, 30-day post training and 60-day post training. There were significant differences in participant scores between the pre-training test and Posttest 1. There were also significant differences on the mean test scores between Posttest 1 and Posttest 2, and 30-day post training and 60-day post training. Table 3 shows a significant effect on tests by time.

Table 3: Individual Analyses of Variance for pre and posttests by group.

		Sum of Squares	df	Mean Square	F	Sig.
Pretest	Between Groups	82.370	2	41.185	1.003	.379
	Within Groups	1191.098	29	41.072		
	Total	1273.469	31			
Posttest 1	Between Groups	394.618	2	197.309	4.502	.020
	Within Groups	1271.101	29	43.831		
	Total	1665.719	31			
30-day	Between Groups	300.618	2	150.309	4.695	.017
	Within Groups	928.351	29	32.012		
	Total	1228.969	31			
60-day	Between Groups	441.643	2	220.821	5.571	.009
	Within Groups	1149.576	29	39.641		
	Total	1591.219	31			

There is a distinct difference in knowledge retention gains between the three groups. The most remarkable difference is the scores of the Profound group on the three posttests compared to the scores of the other two groups on the posttests. Figure 4 provides a graphic representation of the means for the three groups across time.

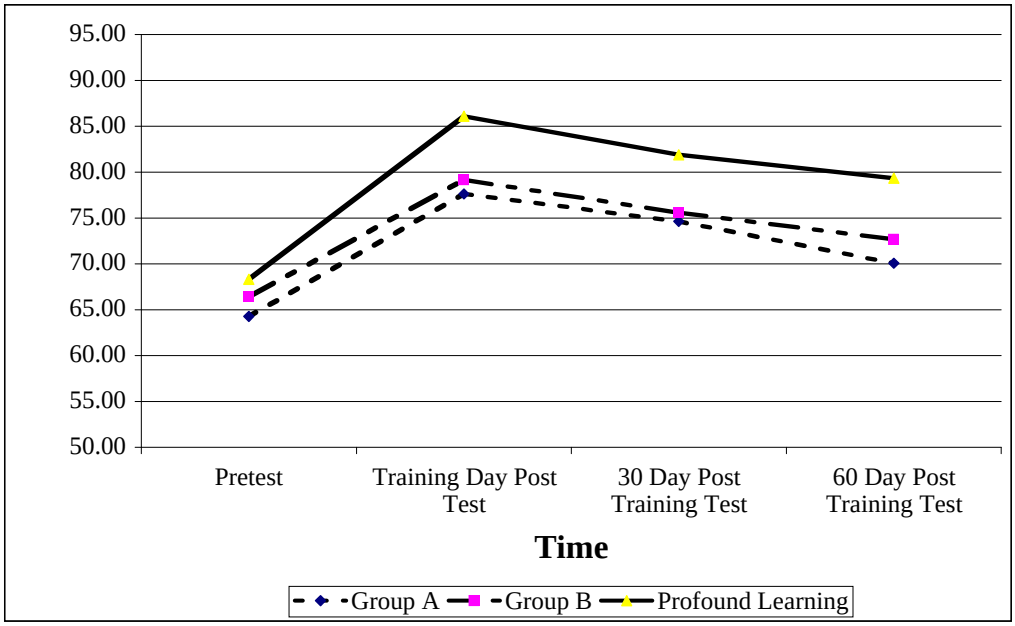

Figure 4: Mean performance of the Focus Q and Profound groups across testing periods.

All participants who started the training, irrespective of group, completed the initial training, but the Profound Learning group took significantly longer to complete the training than either of the Focus Q groups. This appears to be the result of two factors, 1) PLS incorporates review as part of the training, and 2) the delivery medium for training was not the same with both programs. FocusQ had its content on a CD while PLS content was delivered via an Internet connection from their remote server. The Internet connection was sometimes slow in presenting the material so there was a delay in training which occurred for the PLS users. The Profound group had a mean training time of 4 hours, 20

minutes. This was noticeably longer than either of the Focus Q groups. Focus Q without Review had a mean training time of 3 hours, 26 minutes and for Focus Q with Review the mean training time was 3 hours, 20 minutes. The Profound group took, on average, an extra hour to complete the training compared to Focus Q groups. Table 4 presents the means and standard deviations for training time with each group.

Table 4: Means and Standard Deviations for training time with group.

GROUP	Mean (in minutes)	N	Std. Deviation
Focus Q without Review	206.73	11	24.59
Focus Q with Review	200.75	12	46.70
Profound	260.00	9	37.02
Total	219.47	32	44.64

A one-way ANOVA of the time taken to complete the training by group showed a significant difference in training time. A subsequent post hoc test showed a significant difference between the Profound group and the two FocusQ groups.

There was no significant difference between the amount of time taken to complete the training session and the score on the training day posttest. It may be argued that training time was an important factor in achievement on the posttests because the Profound group took much longer to complete the training and scored much better on the training day posttest Conversely, the FocusQ groups took much less time to complete the training and they scored much lower on the posttests. Thus, it might also be argued that the training methodology alone did not affect the achievement on the posttests, but the amount of time taken to complete the training was the important factor in the achievement on the posttests.

The research team was unable to determine precisely if participants were actually reviewing the material over the sixty-day trial period. The data on the participants review rates depended on self-reports. Participants in the Focus Q without Review were not asked to engage in any retention activities and their self-report indicated that none of this group's participants actually reviewed the material during the sixty-day period. Both the Profound and Focus Q with Review groups were asked to review and they reported varying levels of compliance. Table 5 presents a summary of the participant reported review rates.

Table 5: Group percentages and number of participants responding to a question about how often the material was reviewed.

Outlook Study Review Rates by Group

		Reviewed Often	Reviewed Sometimes	Reviewed Rarely
Focus Q with Review		41.67%	33.33%	25%
	n = 12	5	4	3
Profound		66.66%	22.22%	11.11%
	n = 9	6	2	1

Discussion

Significant differences between learners using PLS and both groups of learners using FocusQ were found in this investigation. On the 30 and 60-day posttests, the learners using PLS scored significantly higher on a content test than two groups of learners using FocusQ. There are several possible explanations for the Profound group's strong performance relative to the two FocusQ groups. The scores on the pretest show the Profound group scoring higher before any training was done. Therefore, it might be argued that this group was made up of participants who started the training with some sort of advantage over the other two groups. Time commitments and drop out rates might have been connected in some way. The drop out rate of the Profound group was higher than the FocusQ groups. The Profound group did have to continue with daily log-ins and retention activities, while the FocusQ

groups had much less of a time commitment imposed on them by the study, which may have lead to a higher Profound group dropout rate.

Additional factors may also have affected the results obtained from this investigation. There were a small number of participants (n=32) and high standard deviations for many of the variables. Another factor that might have influenced present findings involves deployment of the two instructional programs. The two CBT programs did not present the content material in exactly the same way during training, therefore this difference might have affected participant results. The FocusQ groups were not able to use the entire FocusQ program because a portion of the program was disabled for this study. FocusQ has built-in unit tests that users are usually required to complete to help them evaluate their learning during training. It was thought these unit tests would give the FocusQ users an unfair training advantage because the PLS system did not have the same unit tests. Also, the Focus Q without Review group had no review activities to do and only had to show up at the evaluation session once a month. This is far less than even the Focus Q with Review group who was asked to independently review the CD-ROM on a regular basis.

The feedback received from each program might have affected the evaluation performances of each group. It has been found that it is very important for learners to get feedback of some kind (Clariana et al., 1991). PLS provided feedback in a very structured way while the FocusQ groups did not have the same type of feedback structure. Logging in and receiving a score everyday from the program might have lead to improved test scores and improved motivation levels in the profound group.

Conclusion

The goal of this investigation was to evaluate the impact of training and the effectiveness of different types of content knowledge retention activities after adult participants used a CBT program to learn about Microsoft Outlook. By comparing three different types of the retention activities the participants completed, it was found that user retention of knowledge could be significantly affected. There were significant differences found in the test scores of the participants in the different training groups. Since the Profound users consistently achieved significantly higher on content tests, it can be argued that there is a relationship between the structure of the instruction and retention activities in PLS and the higher performance on content evaluation tests. The PLS system appears to be a good tool for aiding in self-regulated learning and to be a useful training tool, but more research is needed.

References

Baylor, A., & Chang, S. (2002, October). *Pedagogical agents as scaffolds: The role of feedback timing, number of agents, and adaptive feedback.* Paper presented at the International Conference of the Learning Sciences, Seattle.

Bjorner, J., Kosinski, M., & Ware, J. (2003). Calibration of an item pool for assessing the burden of headaches: An application of item response theory to the Headache Impact Test (HIT). *Quality of Life Research, 12*(8), 913-933.

Boston, C. (2002). The concept of formative assessment. *ERIC Clearinghouse on Assessment and Evaluation*.

Caple, C. (1996). *The effects of spaced practice and spaced review on recall and retention using computer assisted Instruction.* North Carolina State University, Raleigh, NC.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*(6), 633-642.

Clariana, R., Ross, S., & Morrison, G. (1991). The effects of different feedback strategies using computer-administered multiple-choice questions as instruction. *Educational Technology, Research and Development, 39*(2), 5-17.

Collis, B., Boer, W. D., & Slotman, K. (2001). Feedback for web-based assignments. *Journal of Computer Assisted Learning, 17*, 306-313.

Embretson, S. (1996). The new rules of measurement. *Psychological Assessment, 8*(4), 341-349.

Fletcher-Flinn, C., & Gravatt, B. (1999). The efficacy of computer assisted instruction (CAI): A meta-analysis. *Journal of Educational Computing Research, 12*(3), 219-242.

Jackson, S. (2004). Ahead of the curve: Future shifts in higher education. *Educause Review*, *39*(1), 10-18.
Kneebone, R., Scott, W., Darzi, A., & Horrocks, M. (2004). Simulation and clinical practice: strengthening the relationship. *Medical Education*, *38*(10), 1095-1102.
Kulik, J., & Kulik, C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*(1), 79-97.
Lieberman, D., & Linn, M. (1991). Learning to learn revisited: Computers and the development of self-directed learning skills. *Journal of Research on Computing in Education*, *23*(3), 373-395.
Mathan, S. (2004). *Recasting the feedback debate: Benefits of tutoring error detection and correction skills.* Unpublished PhD, Carnegie Mellon University, Pittsburgh, USA.
Okolo, C., & Ferretti, R. (1996). Knowledge acquisition and technology-supported projects in the social studies or students with learning disabilities. *Journal of Special Education Technology*, *13*(2), 91-103.
Outlook. (1997). Redmond, Washington: Microsoft.
Profound Learning Delivery System. (2005). Calgary, Canada: Profound Learning Systems, Inc.
Sinclair, G., Healy, A., & Bourne, L. (1997). The acquisition and long-term retention of temporal, spatial and item information. *Journal of Memory and Language*, *36*, 530-549.
Smith, F. (1998). *The book of learning and forgetting*. New York: Teachers College Press.
Smyth, K. (2004). The benefits of students learning about critical evaluation rather than being summatively judged. *Assessment & Evaluation in Higher Education*, *29*(3), 369-378.
Son, L. (2004). Spacing one's study: Evidence for a metacognitive control strategy. *Journal of Experimental Psychology*, *30*(3), 601–604.
Sprenger, M. (1999). *Learning and memory: The brain in action*. Alexandria, Virginia: Association for Supervision and Curriculum Development.
Streatmans, G., & Eggen, T. (1989). Computerized adaptive testing: What it is and how it works. *Educational Technology*, *38*, 45-52.
Theide, K., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 1024-1037.
Ware, J., Bjorner, J., & Kosinski, M. (2000). Practical implications of item response theory and computerized adaptive testing: A brief summary of ongoing studies of widely used headache impact scales. *Medical Care*, *38*(9), 73-82.

Please cite as: Reid, D., Jacobsen, M. D., Katz, L. (2005). An investigation of computer generated knowledge retention activities in computer-based training with adult learners. *Journal of Learning Design*, 1(1), 77-85. www.jld.qut.edu.au/Vol 1 No 1