Journal of International Information Management

Volume 5 | Issue 2 Article 3

1996

Performance testing with presentation graphics

Floyd Brock
University of Nevada, Las Vegas

Performance testing with presentation graphics

Floyd Brock
University of Nevada - Las Vegas

ABSTRACT

Testing for computer literacy requires extensive administration and controls. The purpose of this research was to investigate the use of computerized, projected quizzes in which students' performances (correct/minute) were measured. As part of these performance quizzes, a series of 16 screens with 48 matching questions were shown for one minute per screen to 64 students. Six weeks later, an improved series of 20 screens with 60 matching questions were shown to 61 students. Last, these two quizzes were compared with a final power quiz with 90 matching questions that was printed on paper. The results showed that students did as well on the second projected, performance quiz as the paper, power quiz. The difference in performance of the genders was unexpected.

INTRODUCTION

Computer fluency, ready computer literacy, is required for success in businesses and government. Sometimes, it is even necessary to qualify as a nerd (Bennett, 1996). To meet these needs and to make their graduates marketable, most colleges require their students to take introductory courses in computers, and some colleges require demonstrated computer competency (CSUSM, 1996). In most business colleges, students first learn personal productivity tools on the computer and later learn how to manage information resources. The sequence assumes some retention and fluency in computer knowledge from course to course and upon graduation.

Demand for these courses and the resulting student enrollment inundate facilities and challenge faculty. More to the point, classes have become quite large; and test preparation, administration, and grading is virtually impossible on a periodic or weekly basis. The purpose of this paper is to explore and to describe a partial solution to this testing problem.

BACKGROUND TO COMPUTER FLUENCY & PERFORMANCE TESTING

When job candidates are asked questions about computers, answers are expected a few seconds later. Their answers must be ready knowledge: computer fluency. Knowledge is not usually determined by hour-long tests. Answering the questions out of sequence or several minutes later is unacceptable, an anachronism of tests printed on paper. Additionally, previous

testing for computer literacy (Brock, 1992) surfaced similar, poignant differences in students other than their scores. Some students finished their tests in much less time than others, and their scores were not necessarily lower than those that stayed longer.

The purpose of this study was to develop a series of quizzes (Quiz I, Quiz II, and Quiz III) that would evaluate computer fluency in a way that better represented the pace of questioning a person would meet during interviews. The first two quizzes were performance (correct per minute) tests, which are similar to a speeded test, " . . . test administered so that students are required to complete the exam within a specific amount of time" (Wiersma, 1990). The difference in these performance tests was that three questions were displayed on screen for a specific amount of time. Also, because these tests do not back up, a person sees these questions only once. The last quiz was a power test (almost unlimited duration), which provided a comparison.

The testing described below evolved from weekly quizzing large freshman classes (60 to 170 students), where duration and control were critical. Similar, updated quizzes that were used with juniors and seniors is the topic of this report. This study opens the plausibility of computerized performance testing in other classes and their continued use as part of students' and employees' evaluations. Similar performance testing could be used to screen job applicants utilizing hundreds at a sitting.

DETERMINANTS OF COMPUTER FLUENCY

This research problem was to determine the computer fluency levels of students in their second Management Information Systems (MIS) course. The term computer fluency levels here means the score on quizzes, a partial knowledge of "hardware, software, systems operations, computer languages, data and information, and systems analysis" (Brock, 1992), in relatively short duration. In other words, computer fluency is computer literacy/time. These quizzes were a three part series that became more comprehensive over the semester. The specific purposes of the study were (1) to develop computer generated, paced projections of questions for performance tests and (2) to find out the efficiencies of the tests. Investigations of different computer graphics, questions, and a few demographic variables was secondary. More important, these computer generated performance tests match the suggestion that " . . . future graphic research should evaluate accuracy and time jointly" (Jarvenpaa, 1989), and appear similar to other computer generated tests (Bell, 1995).

A precondition to the performance tests was that all of the students were exposed to this type of test. All saw a Microsoft Power Point® demonstration in class of a similar quiz administered in the freshman classes. The only essential difference from the demonstration quiz and Quiz I was the content of the questions. Additionally, all received a diskette with the demonstration quiz, so they could view and become accustomed to the testing format. This exposure apparently made no difference (Brock, 1996). The sequence of tests was the demonstration quiz, Quiz I, Quiz II, and Quiz III. The last quiz, at the students' requests, was converted to a power, paper

test, which included a bonus for finishing early. Because of the request, the middle quiz of the series, Quiz II, became the focus of data collection. The numbered quizzes were part of the students' semester grade. To reduce verbiage as with previous computer literacy tests (Brock, 1992), these quizzes contained matching questions of the same format: three items and five options. A few of the questions in the latter quizzes repeated that required knowledge accumulation over the semester. Table 1 shows the differences in the quizzes.

Table 1. Summary of Quiz Characteristics

Questionnaire	Type Test	Media	Semester Week	Chapters Covered	No. Graded Questions	Guessing Penalty
Quiz I	performance	projected	5	1-5	45	none
Quiz II	performance	projected	11	1-11	60	0.25 point
Quiz III	power with bonus	paper	16	1-16	90	0.25 point

The hypothesis of this research study was: performance tests are no less useful than power tests, thus demonstrating the worth of projected questions for a specific duration. The scope included the three quizzes, but specific interests was on the second projected test, Quiz II. It included the most recent improvements and suggestions. Two experimental null hypotheses (H_0) and five additional null hypotheses dealing with demographics were available without notice and with the questions are shown in Table 2.

Table 2. Supporting Research Hypotheses

NO difference in the three quizzes:

$H_0$1: Projected Performance Quizzes (Quiz I & Quiz II)

$H_0$2: Projected Performance and Paper Power Quizzes (Quiz II & Quiz III)

NO difference in performance scores for those students who:

$H_0$3: Are older

$H_0$4: Are of a different gender (specifically, males)

NO difference in the performance scores on questions that:

$H_0$5: Are at the beginning half of the test

$H_0$6: Are more graphic in content

$H_0$7: Have repeating answers

METHOD

Quiz I was administered to two classes during the fifth week of classes in a 60-person, tiered classroom. A ceiling projector, Sony Model VPH 10410, displayed the quiz in 29 *PowerPoint*® slides at a TV resolution on a glass-beaded screen in front of a darkened classroom. During the eleventh week of classes, Quiz II was administered to the same two classes in three separate sessions in a smaller level room. This room discounted parallax and allowed the students to sit where they desired (11 to 28 feet from the screen). An Infocus Systems LitePro 550 displayed the slides at VGA resolution.

The first slides displayed the purpose of the quiz, that it was a performance test, gave the source and chapters of the questions (Stair, 1996), and listed the topics of the questions. All the slides included in this paper were from Quiz II, which had a total of 54 slides. Figure 1 shows the title and topic slides. Question topics were given for context and to prime the students' minds. Two slides showed the instructions, similar to Figure 2; another slide that is not shown gave the hint to "read the answers first." During the display of these beginning slides, the instructor narrated each slide. For example, the instructor reminded the students that the same answer could be used more than once.

Figure 1. Title and Question Topic Slides

(put Fig. 1 here)

Figure 2. Instruction Slides

(put Fig. 2 here)

Shown for one minute each with no narration, the next slides displayed matching questions. The PowerPoint® software provided the seconds that the slide had been displayed in the lower left hand corner of the screen. Figures 3 and 4 are examples of the 60 questions slides, intervening slides that said to look at and darken answers, and the stop slide. The last two slides informed the students that the quiz was over and instructed the students to stand up and to pass their *scantrons* to the center aisle. The instructor narrated these last two slides at full volume and with authority. No time was left for students to scan for and compare answers.

For slide construction, the matching items followed the suggestions of authorities on matching questions (Kubiszyn, 1990) and presentation (Whalen, 1996). Both the short list of numbered descriptions and slightly longer alphabetized list of options were homogeneous; all slides of questions had perceptual proximity, the same format (Wickens, 1995). The topic of the set of questions was at the top of each screen to provide a context for the questions, and all screens with questions were constructed with the same spatial arrangement. In contrast to the random ordering done in some testing programs, the screens were ordered in the same sequence as the chapters of the textbook. The contents of the slides were brief; however, abbreviations and acronyms were explained, as some students requested. Textbook page numbers where placed at the end of each question in brackets so they could be referenced after the quiz.

Quiz III, the paper power test, was administered in a classroom. Students were given one hour and fifty minutes to complete it.

Figure 3. Graphic Questions and Intervening Slides

(Fig. 3 goes here)

Figure 4. Text Questions and Stop Slides

(Fig. 4 here)

RESULTS AND COMMENTS

Table 3 shows a summary of the scores that are normalized to percentages and other differences of the quizzes. In general, the students did somewhat better with Quiz II. The paper quiz, Quiz III, took the students from 16 to 65 minutes to finish. None of the quizzes are statistically independent since knowledge builds through the semester.

Table 3. Summary of Quiz Results

Questionnaire	Duration in min./question	Average % Correct	No. of improved/%	Average Score % (with guessing penalty)
Quiz I	0.33	59.3	---	59.3 (no penalty)
Quiz II	0.34	68.6	49/80%	57.5
Quiz III	0.18-0.72 (0.37 avg.)	65.8	19/31%	57.8

An ANOVA showed that the scores in the quizzes were different, so the scores were tested between the quizzes with t-tests. Quiz II scores, $H_0 1$, were better than Quiz I as would be expected. Quiz II scores also were better than Quiz III scores but not significantly so, $H_0 2$. Older students scored higher than younger students, $H_0 3$, which was not surprising given previous research (Klein, 1993) (Brock, 1992b). The hypotheses, means, results of t-test, and levels of significance (and non-significance, ns) are shown in Table 4.

The fourth hypothesis showing a difference in gender showed the men scored significantly higher than the women; this is a repeat of Quiz I (Brock, 1996). This appears to counter findings on learning in two previous tests on computer literacy (Brock, 1992b); however, the previous two tests did not measure performance. Unfortunately, neither a survey for complaints nor audio recordings about the tests were part of this study. Either may have helped in understanding the gender differences (Kay, 1992), particularly the spontaneous complaints about the stress induced by the tests. The results of Quizzes I and II also agree with other research that found that females do not do as well as males with graphic presentations (Hood, 1993).

Students appear not to have tired as the Quiz II progressed, $H_0 5$; this was unexpected, particularly for two night classes and how they faded on Quiz I. The 12 questions embedded in graphics, $H_0 6$, perhaps were too few to be compared with the 48 questions in text. Without three graphics questions that were identical to three questions in the previous quiz, the percentages changed significantly in favor of text. Compared with questions with single, unique answers, questions with the same repeating responses were missed significantly more often. Guessing perhaps showed in $H_0 7$, and on reflection, this performance test needs a correction factor for wild guessing (Popham, 1990).

Table 4. Results of Hypotheses Tests

HYP	Independent Variables	N	Mean %	Std Error	t-Score	Signi-ficance
$H_0$1:	Projected					
	Quiz I	64	59.3	12.6	-4.31	*p<0.05*
	Quiz II	61	68.6	11.4		
$H_0$2:	Projected and Paper					
	Quiz II	61	68.3	11.4	-1.40	*ns*
	Quiz III	62	65.8	10.9		
$H_0$3:	Median Age (24.1)					
	Older than	30	71.4	10.2	-1.90	*P<0.05*
	Younger than	30	65.9	12.1		
$H_0$4:	Different gender					
	Males	32	71.1	11.7	1.84	*P<0.05*
	Females	29	65.9	10.6		
$H_0$5:	Question placement					
	First 30	30	62.8	23.6	-2.03	*P<0.05*
	Second 30	30	74.3	19.9		
$H_0$6:	Content more:					
	Graphic	12	59.2	23.8	-1.56	*ns**
	Characters	48	71.0	21.6		
$H_0$7:	Answers that:					
	Repeat	9	53.0	31.3	-1.70	*ns**
	Not repeat, unique	51	71.4	19.6		
						*P≈0.06

CONCLUSIONS

Projected performance testing works as well as paper power testing. This was a delightful surprise that supports Bugbee's research (Bugbee, 1996). When the tests are constructed off site and administered only one time, the security of the tests can be controlled absolutely. Test administration and recapture take minutes. Performance testing severely limits time for educated guessing by elimination and proximity scanning for other answers. Because paper and printing are reduced, projected testing can unburden both the student and evaluators. From observation, computerized projection of tests in the classroom appear to work as well as those computer administered in the laboratory filled with uncontrolled distractions and onlookers.

Students do, however, express concerns about the stress the performance tests induce. These concerns were so great that to quiet their discontent, the third quiz was converted to a two-hour, paper quiz. Complaints may be further evidence of "worry often ha[ving] detrimental effects on performance" (Eysenck, 1995). Students similarly expressed concern over an oral exam. Audio recordings of their spontaneous complaints and concerns may produce a rich source for future research. They appeared to want a tactile relationship with questions, " . . . a proper test," as one student recently commented. On the other hand, in listening to their comments over a semester, I believe that their uncorrected visual deficiencies may be a factor that needs to be considered for visuals presented in the classroom. Many just cannot see well.

On this examination of projected testing with time constraints, both screen and classroom design issues did arise. Given a short duration, say a minute, how many characters on a screen are too many to read? Does the rearrangement of graphic elements from those seen in the text complicate the questions? How can the limitations of software and equipment, such as screen resolution and size, be worked around? This study on performance testing with computer projections is just beginning.

REFERENCES

Bell, P. C. & O'Keefe, R. M. (1995). An experimental investigation into the efficacy of visual interactive simulation. *Management Science, 41*(6), 1018-1038.

Bennett, J. (1996). *The [infamous] nerdity test* [Online]. Available World Wide Web: http://gonzo.tamu.edu/nerd.html.

Brock, F. J., Kohl, J. P., & Thomsen, W. E. (1992). Questions on computer literacy in the first computer class. *Journal of Research on Computing in Education, 24*(4), 563-570.

Brock, F. J., Kohl, J. P., & Thomsen, W. E. (1992b). Improvements in computer literacy linked to gender and learning style. *Journal of International Information Management, 1*(1), 25-40.

Brock, F. J. (1996). Performance testing for computer literacy. *Proceedings, Educational Studies and Research, The International Association of Management, 14*(1), 31-36.

Bugbee, A. C. (1996). The equivalence of paper and pencil and computer-based testing. *Journal of Research on Computing in Eudcation, 28*(3), 282-299.

CSUSM [California State University San Marcos] Class Schedule. (1996). Computer competency requirement [online]. Available World Wide Web: http://144.37.1.41/cwis/class_schedule_f96/comput_comp_req.html.

Eysenck, M. W. & Keane, M. T. (1995). *Cognitive psychology,* (3rd Ed.), Lawrence Erlbaum Associates, 457.

Getting Results with Microsoft Office for Windows 95 (1995), 255-291. Redmond, WA: Microsoft Corporation.

Grace, R., Darnel, R., Stevenson, N. & Toliver, P. (1995). *Special edition using PowerPoint for Windows 95.* Indianapolis, IN: Que Corporation.

Hanna, G. S. (1993). *Better teaching through better measurement,* 267-271. Fort Worth, TX: Harcourt Brace Jovanovich, Inc.

Hood, J. N. & Togo, D. F. (1993). Gender effects of graphics presentations. *Journal of Research on Computing in Education, 26*(2), 176-184.

Jarvenpaa, S. L. (1989). Task demands and graphical format in information processing. *Management Science, 35*(3), 285-303.

Johnson, S. M. (1994). *Microsoft PowerPoint 4 for Windows step by step.* Redmond, WA: Microsoft Corporations and Prespection, Inc.

Kay, R. H. (1992). Understanding gender differences in computer attitude, aptitude, and use: An invitation to build theory. *Journal of Research on Computers in Education, 25*(2), 159-171.

Klein, J. D., Knupfer, N. N. & Crooks, S. M. Differences in computer attitudes and performance among re-entry and traditional college students. *Journal of Research on Computers in Education, 25*(4), 498-505.

Kubiszyn, T. & Borich, G. (1990). *Educational testing and measurement* (3rd ed.), p. 74-77. Glenview, IL: Scott, Foresman, and Company.

Popham, W. J. (1990). *Modern educational measurement: A practitioner's perspective* (2nd ed.), p. 336. Englewood, NJ: Prentice Hall.

Stair, R. M. (1996). *Principles of information systems: A managerial approach* (2nd ed.), p. 2-191. Danvers, MA: Boyd & Fraser.

Whalen, D. F. (1996). *I see what you mean: Persuasive business communication*, p. 190-199. Thousand Oaks, CA: SAGE Publications, Inc.

Wickens, C. D. & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundations and relevance to display design. *Human Factors, 37*(3), 473-494.

Wiersma, W. & Jurs, S. G. (1990). *Educational measurement and testing* (2nd ed.), p. 404. Needham Heights, MA: Allyn & Bacon.