Perceived Performance in Internet Testing

Wm. Benjamin Martz, Jr.
wmartz@uccs.edu

Morgan M. Shepherd
mshepher@uccs.edu

both from:
Information Systems Department
College of Business
University of Colorado at Colorado Springs
1420 Austin Bluffs Parkway
Colorado Springs, Colorado 80933-7150

Abstract

With the growth of distance education, many issues pertaining to how students perform in this environment are surfacing. The differences between the environments may have pedagogical implications on the basic activities in education such as assessment. This paper provides the results of comparing the testing process between two sections of a graduate programming class, where one was an on-campus class and the other a distance class. The course content, instructor, syllabus, lecture materials, notes, assessments and semester (time of year) were the same. Both groups were surveyed to test their satisfaction with the testing procedure and with their perception of certain aspects of the social environment. While there was no difference in actual test performance, the results showed differences in perceived test performance. Two conjectures about possible causes underlying the difference and suggestions for possible future research end the discussion.

1.0 Introduction

Distance education is big business; over 1.6 million students enroll in distance education classes annually. The Institute of Higher Education [17] predicts that by 2004 over 90% of all two and four-year colleges will offer some sort of online courses. By the end of 2003, the virtual education market will grow in excess of $21 billion [26]. New technologies and improvements in networking capabilities are enabling distance education instructors to come closer to providing the traditional learning environment for their students. However, there are still many issues that need to be resolved, and according to recent research studies it does not appear that these will be resolved anytime soon [2].

Interestingly, the research in distance education goes back into the late 20's when the first studies were published comparing the test scores of students in a classroom to their counterparts in a correspondence course [9]. Since then, hundreds of journal articles, studies and reports have been published with similar comparisons with TV, radio, video tapes, computer-based training, audio-conferencing, groupware, and now the internet, representing the technology compared to the traditional classroom [4], [21].

Much of the research compares the distance education approach to the traditional classroom approach, looking for areas where the results from the distance education approach equals or exceeds those from the traditional classroom approach. There are many confounds involved, making this type of research difficult. One of these confounds is the definition given to “distance education.” These definitions range from correspondence courses, to satellite classrooms where the instructor travels to lecture to a group of students meeting face-to-face, to courses that are held via email, to courses that are held via two-way full motion video with other technological support. The definition debate will probably never end, as some definitions do not include the use of any technology (correspondence courses),
while others require several technologies to be implemented (full motion video with chat rooms, listservers and email). For the purpose of this study, our definition of distance education involves a student body who never see each other or the instructor; who communicate via email, phone or chat; and who hand in assignments via email or via posting to a common work group area.

Other confounds arise due to the nature of the dependent variables that are studied, and the interaction affects between them. Some of the dependent variables studied have been student performance, student satisfaction, and student retention. Within each of these three variables, socialization is thought to play a significant role. For example, Kling [19] defines the study of complex issues around the interaction of information technologies with an institutional and cultural perspective as “social informatics.” With the proliferation of the Internet, distance education is becoming a legitimate arena for social informatics. Kling’s definition guides us to look at social issues brought on by the use of the Internet technology with education in general.

Other researchers ([13], [23], [25]) suggest that learning requires a social context to be effective. Bandura [3] relies heavily on the interaction of people with their environment in his Social Learning Theory. Gunawardena and Little [11] use results from their study to conjecture that “social presence is a strong predictor of satisfaction” in a computer-based environment. Tu [29] goes even further by suggesting that the main driver of learning is the “consciousness of another person in the environment.” Each of these studies indicates the importance of social issues in distance education. However, for each study that claims to show a significant difference between the distance and the traditional classroom environment, there seems to be one that claims to show that there is NO significant difference between the two environments. TeleEducation NB [28] has helped compiled two lists of research with regard to distance education. Russell [24] summarizes those studies that found “no significant difference” between students in distance education and students in traditional classroom environments. Those findings demonstrating a significant difference are available through the TeleEducation [28] website.

Some studies have found significant differences in favor of the distance learning environment. For example, in a study on participation, Colorito [8] found online students participating more than traditional students. Neslar and Hanner [22] surveyed students from different nursing programs about their level of socialization and were surprised to find online nursing students showed more socialization characteristics than their peers in the traditional classroom environment. Gagne and Shepherd [10] found online students less satisfied than their peers in the traditional classroom with the availability of the instructor in a distance environment. However, this does reinforce the importance of some type of social interaction to distance education students. Maki [20] found satisfaction with a lecture course to be lower for online students than for traditional students. Chen [7] found that the attitudes of students in a computer-based class were less positive than those attitudes from students in a conventional classroom.

By definition, much of distance education is a solitary experience. Hogan and Kwiatkowski [15] argue that the emotional aspects of teaching large groups with technology have been ignored. Hearn and Scott [14] concur and suggest that before adopting technology for distance teaching, that the technology must be able to address the social context of learning. This idea is somewhat borne out by Hill’s et al [13] work with Computer Based Training (CBT) wherein they found that students were more successful in the CBT environment when more social interaction occurred.

Several researchers suggest that there may be problems inherent with the distance environment. The most bothersome may be student retention where course drop out rates as high as 50% are reported [6]. Brown and Liedholm [5] found that students in the virtual classroom performed worse than their traditional counterparts on examinations. Without the hands-on supervision inherent in the traditional environment, cheating and plagiarism take on added concerns [1].

Other social factors also are being examined. Kirkman [18] identified two social factors, creating trust and the feeling of isolation, as key factors for students to succeed in the distance environment. Horn and Kling [16] research and report on the concern that distance courses may create new anxiety and stress for the student. Haythornthwaite [12] found that the distance environment limited the number of social cues and in turn, this limitation reduced the participation found in a distance environment.

This study looks more closely at student satisfaction, specifically student satisfaction with the testing process in our distance education program. To help reduce the threat of cheating, all student tests are timed. A timer starts when the student begins the test. The timer counts down the time, and is clearly visible in the corner of the screen. After the timer is started, the student can view the questions and take the test. Should the timer expire, the students’ work up to that point is automatically submitted.
One potential problem with this method is the increased stress that might arise. Building on the work of Hora and Kling [16] and Haythornthwaite [12], we hypothesize that this stress comes from two sources. The first source is the fear of not performing well, and the second source is the fear of not performing as well as your peers. Both fears exist in a traditional classroom, although the latter fear may be somewhat lessened due to the fact that the student can receive some visual clues as to how well they are doing compared to their peers.

In the traditional test-taking environment students get feedback from the other students in the room. This feedback takes several forms. The first is in the form of pre-quiz, “working” of the room. Students ask each other how much they studied for this quiz and how difficult/easy they think it will be. A second form of feedback occurs during the actual taking of the test. In the role of a student, suppose I see that time is running out and I’m only 75% of the way through the quiz. If I notice that 90% of my peers are still taking the quiz I may realize that at least we are all “tanking” this quiz together (misery loves company). This socialization cue will help to dampen my fear. Yet another cue is in the form of the post-quiz “working” of the room. Students get a chance to see how well the others thought they did (albeit perceived performance) on the quiz. The distance student does not have easy access to any of these socialization cues. Hence we decided to explore the effects of these socialization aspects on student satisfaction in the distance test-taking environment.

We hypothesized that due to this lack of socialization cues in the distance education environment, the distance students would be less satisfied with their performance on the quiz than the traditional classroom students. We devised the following experiment to test this hypothesis.

2.0 The research study

Like many of the other studies discussed earlier, this research study was undertaken to explore the possibility of differences between the traditional classroom environment and the distance education environment (Figure 1). However, the unit of analysis for this study is the common learning activity of test taking. Most studies noted earlier use the complete course as the unit of analysis i.e. student performance over the span of an entire course. This study uses a subset activity of a course, a test. The rationale of using this lower level of granularity is that differences that may occur at the test level may well be “washed out” when the aggregate unit of measure, the course, is used.

The subjects in this study were drawn from two Masters of Business Administration (MBA) level classes. Both classes were introductory, information systems classes teaching Visual Basic© programming. One was offered through an established distance program while the second was taught in a traditional classroom. Both courses were offered in the same 16 week semester format, by the same instructor, using the same book, same lecture material including slides, same assignments and the same grading scale. The weekly assignments and order of topics was the same. The distance class had 9 students and the on-campus class had 10 students.

The actual test used in the study was the first test of the semester for both classes. The distance class took the test using an automated, online testing program provided by the course content package that was accessed over the internet. The on-campus class took
its test on paper. The tests had the same questions with the same weights and the same order. The test was divided into three sections so as to minimize the potential problems in the distance environment. Both environments imposed the same time limit; the online test automatically enforced the time limit while a timer was used to enforce the on-campus class time limit.

A 15-question, questionnaire was developed using the themes of interest from the concerns and issues raised from the previous distance education research cited earlier. Questions were developed to explore the areas assumed to be fundamentally different between distance and traditional classroom education. The questions were developed to look for potential differences in perceived performance, test satisfaction, study habits, and time pressure.

In both cases, immediately after the test, the questionnaire was distributed to the class members. The distance class received the questionnaire as an email attachment. All questionnaires were returned by students prior to receiving their grades for the test. The on-campus class filled out and turned in the questionnaire before leaving the classroom. All students in both classes returned their questionnaires.

### 3.0 Results

The questionnaire responses along with the actual test scores from the two groups and were tabulated and are shown in Table 1. Students were asked simply to agree (Agree) or to disagree (Dis.) with the prompting statements on a scale of 1 (Low) to 10 (High). Table 1 shows the significance of a statistical test, the Fisher Exact Test (Siegel, 1958), comparing the results of the online students to the on-campus students for each question.

#### Table 1: Comparison of online to on-campus groups

<table>
<thead>
<tr>
<th>Statement</th>
<th>Online</th>
<th>On-Campus</th>
<th>Sign. (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 I felt a lot of time pressure in taking this test.</td>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Q2 The time pressure helped me to perform well on this test.</td>
<td>0</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Q3 I would have scored higher with more time for the test.</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Q4 I prefer taking tests with my other classmates in a traditional classroom.</td>
<td>0</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Q5 I prefer taking tests by myself w/o other classmates in the room.</td>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Q6 The test was a fair way to evaluate how well I knew the subject matter.</td>
<td>6</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Q7 I studied for this test, more than I usually do for tests</td>
<td>0</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Q8 I studied with others (study group, non-classmate, etc.) for this test.</td>
<td>1</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Q9 I prefer studying with others for tests.</td>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Q10 I believe that I finished sooner than 50% of the others taking this test.</td>
<td>2</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Q11 I believe that I scored better than 50% of the others taking this test.</td>
<td>2</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Q12 Overall, I was satisfied with my performance on the test.</td>
<td>2</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Q13 Overall, I was satisfied with the format of the test.</td>
<td>6</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Q14 I waited until the last possible time to turn my test in.</td>
<td>3</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Q15 I studied between the sections of the test.</td>
<td>2</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

Non-Questionnaire Objective Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Online</th>
<th>On-Campus</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Prediction of Position (CPP)</td>
<td>5</td>
<td>8</td>
<td>.259</td>
</tr>
<tr>
<td>Correct Prediction of Score (CPS)</td>
<td>5</td>
<td>4</td>
<td>.414</td>
</tr>
<tr>
<td>Actual Test Score - ATS (average)</td>
<td>127</td>
<td>133</td>
<td>.211(2)</td>
</tr>
</tbody>
</table>

(1) Fisher’s Exact Test (2) Mann-Whitney Test - [27]

The first important observation is that two of the questions (Q4 and Q12) show a significant difference. Q4 hints at a social interaction difference between the two groups. Clearly, online students do not want to take tests with other students around while the on-campus students seem to prefer to have other students around.

Question 12 proves a little more interesting. Q12 shows that the two groups of MBA students report a significant difference in their satisfaction with their own performance. Remember, this is really perceived
performance, as the results were not yet returned in class or available online. This finding becomes more noteworthy because the actual test scores were not significantly different.

Questions 1, 2, and 3 deal with the issue of time pressure. Did the two groups feel pressure to complete the test? No, not really. Both groups disagreed that they felt time pressure and they disagreed that the time pressure helped their performance. While not statistically significant, the results for Q3 show that more of the online students would have preferred more time. Combined, these results show that a time limit did not induce the feelings of pressure, but the online group would better receive an extension of the time limit.

Questions 4 and 5 tried to determine differences in the social aspect of having others in the room when taking the test. Certainly, being able to see when others complete and hand in the test could induce comparisons. In turn, these comparisons, if negative, could induce higher levels of anxiety. The results of these questions track well with what would be expected. Those students that have chosen a distance education class report little interest in having other students as a reference group and those choosing an on-campus environment like having other students as reference.

Question 6 asks directly for an assessment of the fairness of the test. Both groups thought the test was a fair evaluation. This can be interpreted to mean that neither the technology nor the test seemed to impact on the students’ perception of fairness.

Questions 7, 8 and 9 concentrate on the students’ study habits. Questions 8 and 9 try to determine if there is a difference in a student’s preference for studying with others. There was none. Question 7 was asked to see if there was any difference in perceived difficulty of the test. Neither group studied more than they normally do. This is significant for the online group as it implies that they did not envision the distance exam to be harder than what they were used to.

Questions 10 and 11 were designed to test for the potential social impacts of the on-campus environment. As suspected, the on-campus students did better than the distance students when they were asked to predict in which half, top half or bottom half, of the class 1) they had completed and handed in their test, and 2) they had scored. The on-campus students were very accurate with their assessment of whether they turned their test in with the upper 50th percentile, or the lower 50th percentile. Half estimated that they finished in the first 50th percentile, and half estimated that they finished in the second 50th percentile. It is also interesting to note that their estimate as to whether they scored in the upper 50th percentile or the lower 50th percentile was 50-50.

Questions 12 and 13 deal with a student’s satisfaction. The first question looks at his or her performance and the second asks about the format of the test. As stated earlier, the perceived performance proved significantly different between those students in the on-campus and those in the online class. This difference does not carry through on Q13 as no difference was found related to the format of the test.

Questions 14 and 15 compare each student’s behavior with respect to when they handed in the test and to whether or not they studied between the sections of the test. One could expect the on-campus students to wait until the last moment while the distance students do not. The results imply otherwise. Question 15 is truly unique to the distance environment. It indicates that two students found that it was necessary to study between the sections of the test.

4.0 Discussion

The first major limitation of this study is the sample size. With only 19 subjects spread over two classes, any differences found between the two groups will be sensitive to statistical prejudices. It should be noted that the Fisher Exact Test is the appropriate test for analyzing a two by two result for significant differences. Also, it is a very conservative test meaning that if a significant difference is noted, it is more likely a true difference.

In the counterargument, one that argues for the reliability of the results, many of the typical confounds are not found in this study. For example, all subjects in both classes returned the survey. The results represent the whole population and not just a sample. Additionally, the courses were essentially the same except for the independent variable, the medium; the instructor was the same, the MS PowerPoint© and lecture notes were the same, the grading was the same and the test questions were the same format and in the same order.

Notwithstanding the issues of a low sample size, the results still provide some interesting discussion points. First, why would there be a difference in perceived performance? A further analysis of the data did not find any of the other questions (Q1 through Q15) significantly correlated to Q12. So, none of the typical concerns proposed such as “time pressure (Q1-Q3),” “fairness of the test (Q6),” “study habits (Q7-Q9),” nor “format of the test (Q13),” help us understand the perceived performance difference exhibited.
However, one interpretation of Q10 and Q11 may provide a hint. While the numerical differences on these questions between the subjects in these two classes is not statistically significant, it is interesting to note that the estimates of finishing position in the on-campus class better match the actual results than do the estimates in the distance class. This hints that the students in the on-campus class have a better reference point. Obviously, they can see the other students handing in their exams and compare their relative position. This reference point was not available to the distance class. Could it be that the reference point provided in the traditional classroom helps students understand their relative position and lead to better “perceived satisfaction”? This might also explain the decidedly negative impression that the distance students felt, concerning both whether they turned in their test in the upper or lower 50th percentile, and whether they scored in the upper or lower 50th percentile.

The second area of interest is with time pressure. Again, without statistical significance, this is conjecture only. The numbers for the time pressure questions hint at time pressure impacting the online students more than the on-campus. This is especially true when a student was asked if they would have scored higher with more time (Q3). The two groups had the same amount of time. But we need to realize that a key difference was the “enforcer” of the deadline. In the distance case, the software counts down using a displayed clock and when zero is reached, the software closes down the test and locks the student out. This is a very rigid and impersonal activity in that there is not arguing with the computer. In the classroom environment, the “enforcer” is a human being. A student may feel it easier to push the time limit. Even as the instructor is saying “It is time to turn the tests in,” a student may feel comfortable finishing a sentence. Extrapolating then, could it be that the “impersonal ness” of the automated testing environment influences perceived pressure.

The results raise one concern and provide two conjectures for understanding differences between on-campus and distance environments. The concern is for why would online students report lower perceived performance than on-campus students. Several keys factors suggested by the literature such as test fairness and study habits did not help explain the difference. However, one conjecture consistent with the data is that an on-campus student has a built-in reference group when handing in tests; those students that have already handed the test in before them. The second conjecture consistent with the data is that “time pressure” may be inherent to the distance testing environment simply because the testing software is an impersonal enforcer of the time limit.

Lack of social cues may very well account for the more pessimistic attitudes of the distance students, even though there was no significant difference on the test. Now we need to look at various ways in which we can infuse the distance education environment with similar social cues. One way may be to put an “average time to complete quiz” number up on the screen for the distance students when they start their quiz, along with a “you are the XXth person to take this test” number so the students can better make sense out of the two. We may need to hold a special chat session prior to the quiz to allow the distance students to get some of these socialization cues. If we can infuse some portion of these social cues into the distance environment, we may be able to improve the satisfaction levels of the distance education students taking timed quizzes.

5.0 Conclusion

The distance education environment is big business. As such, there is much interest in understanding any differences inherent between these environments. The literature has a long list of researchers concerned with understanding these differences. Many hard, objective issues have been raised such as performance, retention and participation. Lately, softer, subjective social issues such as time pressure, anxiety, stress, and trust have been thrown into the debate. All in all, the results have been ambiguous. There are studies showing significant differences between the two environments and there are studies showing no significant differences. The later imply that distance environment is “just as good” as the traditional environment.

The study reported here adds one more issue to the significant difference side of the equation. Using two classes, equivalent on many of the confounding variables such as instructor, test format, time constraints, etc., data was gathered on issues such as perceived performance, study habits, and actual performance. The limitations of a small sample size notwithstanding, the results show that the distance students rated their perceived performance significantly lower than the on-campus students.

The attempt to interpret the implications of these findings suggested two conjectures consistent with the data. One, the testing process in the distance environment uses more of an impersonal enforcer of the time limit. As such, the distance student may feel more time is needed and less positive about his or her performance as the time limit runs out. Two, the on-campus environment provides students with a built-in
reference group for analyzing their performance. As students hand in tests, they “know” where they fall in relation to others that have and have not yet handed in the test. Possible methods for adding the reference group to distance classes and for removing the personal confound from the on-campus environment were suggested. The results of this study suggest one more area of concern for those looking to implement a distance environment: Students in distance environments may exhibit lower perceived performance.

6.0 References


[17] Institute for Higher Education Policy. (2000). Quality on the line: Benchmarks for Success in Internet Distance Education. Washington, D.C.


