

A Statistical Analysis of Student Performance in Online Computer Science Courses

Merry McDonald

Northwest Missouri State University
Department of CSIS
Maryville, MO 64468
00-1-660-562-1136

merry@mail.nwmissouri.edu

Brian Dorn

Iowa State University
Department of Computer Science
Ames, IA 50011
00-1-515-294-3477

dorn@cs.iastate.edu

Gary McDonald

Northwest Missouri State University
Department of CSIS
Maryville, MO 64468
00-1-660-562-1199

gary@mail.nwmissouri.edu

ABSTRACT

During the past few years the number of courses offered online has greatly increased as technology has made delivery of such courses feasible. This paper discusses the experiences of the authors in teaching online courses for upper-division computer science students and reports results of a study comparing success of students in online courses to students in traditional courses.

Categories and Subject Descriptors

K.3.1 and K.3.2 [Computers and Education]: Computer Uses in Education - *Distance Learning*; Computer and Information Science Education - *Computer Science Education*.

General Terms: Experimentation, Human Factors

Keywords: Online learning; Distance learning; Course delivery methodology.

1. INTRODUCTION

During the past few years the number of courses offered online has greatly increased as technology has made delivery of such courses feasible. In many cases, entire degree programs are offered online, many of which are in Computer Science or Computer Information Systems. The University of Phoenix offers online degrees, several of which are in technology-related areas [11]. MIT's OpenCourseWare program makes course materials available online to the general public [9]. A quick web search using terms like "computer science online degree" will yield virtually thousands of results. At Northwest Missouri State University, two of the authors have been heavily involved in putting all required courses for the Computer Science major online. Developing online courses for computer science students has special challenges that are specific to the discipline. This paper discusses the authors' experiences in developing and offering such courses and reports results of a study comparing success of students in online courses to students in traditional courses. While much of the information applies to online courses in any discipline, the focus of the paper will be online courses in computer science, with special attention to upper-division courses.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGCSE '04, March 3-7, 2004, Norfolk, Virginia, USA.
Copyright 2004 ACM 1-58113-798-2/04/0003...\$5.00.

2. TERMINOLOGY

Throughout this paper, we use the term *online* to refer to courses offered via the web, with no required face-to-face contact with the instructor. The term *onground* refers to courses taught in a traditional classroom environment.

3. WHY OFFER ONLINE COURSES?

Online courses are attractive to a significant number of students and faculty because of the flexibility that such courses offer. Students have busy schedules. Some students work many hours per week and need to have more flexibility in planning their study schedule. Other students live far enough from campus that driving in each day requires a significant amount of time. Students with special needs, such as single parents, appreciate being able to arrange their study time around other demands over which they have little control. Students participating in university-sponsored extracurricular activities are sometimes required to be absent from campus for considerable periods of time, or cannot take courses offered at certain times of the day.

In the case of computer science students, online courses offer an opportunity to learn important skills that will be of help to them in their future careers. Employers at the authors' institution repeatedly emphasize the need for independent thinkers who can learn new material on their own. No institution can offer a computer science education that provides the detailed vendor-specific knowledge that a student needs on a job. However, the institution can provide a firm foundation of theoretical and practical knowledge, on which the student can build the additional skills necessary for success in a career. In addition, the institution can ensure that students learn how to learn. As they progress in their careers, they will be constantly required to learn new skill sets. In most cases, formal classes will not be provided for them. Instead, the employee will be expected to learn the new skills using web and printed resources. Many companies are using Learning Management Systems as a central component of their corporate training program. [3] Learning Management Systems are web-based programs that offer a variety of tools for delivery of content and performance analysis. It is especially important for computer science students to be able to use such systems effectively, because they will be working in a field which is changing rapidly and necessitates constant updating of skills and knowledge.

Online courses provide more choices for the student. As more complete degree programs are offered online, location of the student and the faculty member, and the existence of a university campus in the traditional sense, will become increasingly irrelevant. This increase in choices will extend even to traditional students on a residential campus. Most traditional students may continue to prefer

the social setting of a residential college, but may choose to take some individual courses from an online, possibly different, provider.

4. ONLINE COURSES FOR COMPUTER SCIENCE STUDENTS

In some ways, offering online courses for computer science students is easier than for other students. Probably the biggest advantage is that most computer science students are comfortable with computers and web sites. In other disciplines, a student may avoid online courses for no other reason than fear of technology. In fall 1999, Kleinman and Entin [5] compared performance of online and onground students in an introductory computer science course. This course was taken by students in a variety of disciplines. A significant number of the online students withdrew, many at the beginning of the semester, apparently because of the technology hurdle. The authors of this paper have had extensive experience teaching online courses to computer science students, and in our experience, there is no technology hurdle for this group of students.

Some might also argue that another advantage is that computer science instruction relies less heavily upon in-class discussion than some other disciplines. As in most technical and scientific fields, there is more emphasis on mastering the fundamental concepts, theories, and basic practical skills, rather than on debating fine points with other class members. This is true to some extent, but certainly many fine points of the concepts and theories are covered in the class lecture and in response to questions that come up in class. For example, in our upper-division Database Systems class, database design is covered extensively. In designing a database, many issues arise, not all of which can easily be covered in formal lecture notes. Some issues are quite subtle, and are more effectively introduced through in-class work where students have a chance to wrestle with the issue before it is addressed formally by the instructor. These class exercises can be provided to online students, with a discussion of the solution, including the subtle points discussed in class, but the question remains as to whether this is as effective as having students first think through it themselves.

Another issue that arises in the development of online courses for computer science majors is how to best help the student master the necessary technical knowledge. A few students can learn a great deal from reading a textbook and experimenting on their own, but most students need much more guidance in order to be successful in any course, and especially in an online course. In order to master the necessary knowledge, students need to be provided with materials that will aid in the learning process. These materials may include notes that summarize the major points, sets of examples that further illustrate the concepts, and many sets of exercises for students to practice the basic skills that they are attempting to master. Development of these materials is time-consuming for the instructor, and students must be proficient readers and sufficiently motivated to complete the exercise sets in a timely manner. The authors provide materials such as the ones described here for all students, not just online students, and have done so for many years, before online courses were offered at all. However, in preparing materials for traditional lecture-based courses, most professors rely heavily on the fact that they will be presenting a lecture based on the notes, and that they will see students on a regular basis, making it easy for the students to ask questions about any exercise sets. Materials for online students must be much more detailed, and there is a greater burden on the professor to produce error-free materials. In computer science, this can be especially problematic. By the time a good set of materials has been developed and refined over a period of

two or three semesters, it is frequently necessary to make significant modifications due to changes in the field.

One might expect that computer science students taking courses online would be eager and willing to use the many forms of electronic communication available through the course websites. However, in spite of the fact that computer science students feel comfortable with the technology, the authors have observed in their online classes that traditional, residential students who may be taking only a few courses online still want face-to-face contact with the instructor. The authors allow their online students to visit them during office hours, and help sessions are frequently set up for all students, online or onground. Many of the online students take advantage of these services and also visit the class assistant during his/her lab hours. It is also interesting to note that, although computer science students are adept at using the technology and have extensive experience using electronic communication, such as chat rooms, the authors' experiences indicate that students are reluctant to use discussion threads to collaborate with one another. When team projects are required, the students report that the biggest problem is finding a common meeting time, and yet they resist the added flexibility that electronic collaboration, including asynchronous models, might afford. Kleinman and Entin [5] reported a similar finding regarding this issue.

5. ONLINE VS. ONGROUND COURSES

Exactly how do online and onground courses differ? For the courses taught by the authors, the similarities and differences are outlined below:

- the same content is covered in both onground and online courses
- PowerPoint slides used in lectures are available to both online and onground students
- all worksheets and exercises are available to both online and onground students
- students in onground and online classes have the same major assignments
- students in online and onground sections take the same exams
- all exams are supervised
- onground students attend class regularly; online students do not
- onground students may have some in-class activities that are not duplicated for online students – for example, weekly quizzes may be given to onground students but not online students
- both online and onground students may see the instructor during office hours
- out-of-class help or review sessions conducted by a lab assistant may be attended by both online and onground students
- online students who cannot easily come to campus can arrange for one-on-one help from the instructor or lab assistant, either by requesting that a private chat room be set up, or arranging for a telephone visit
- both online and onground students can use discussion threads on the website to post questions about any material covered in the course, including exercises and assignments
- e-mail is discouraged for both onground and online students, except for private correspondence; both groups of students are encouraged to use the discussion threads to post questions whenever possible.

6. PERFORMANCE ANALYSIS

As with any new teaching practice, it is necessary to step back and evaluate whether or not it actually works. As early as 1997, educators were evaluating online instruction [7]. Of the more recent

work that has been done in this area, only a few papers include a statistically rigorous analysis. Kleinman and Entin [5] compared in-class and distance-learning students in terms of both performance and attitude. Olson [10] compared online and lecture methods for delivery of the CS1 course. Korhonen et.al. [6] conducted a study comparing performance of students instructed during classroom sessions versus students using a web-based learning environment. These studies varied in methodologies, type of courses being studied, types of students involved in the study, and in the component chosen to measure of success – grade in the course, grades on exams, or simply completion of courses. All three studies were for a single semester. Also, all three concluded that there was no difference in online versus onground instruction.

During the 2001-2002 academic year, the authors began a study with the aim to provide insight into the effectiveness of online teaching specifically in upper level Computer Science courses. This initial study [8] focused on student performance in two undergraduate courses: Computer Networks and Database Systems. Analysis of this data showed that students in the onground networking section significantly outperformed the online students in the same course. However, students in the onground database course did not exhibit a statistically significant advantage over the online students. These mixed results based on only one academic term led to a long-range analysis, the results of which follow.

6.1 Approach

Since the fall of 2001 we have gathered data from both online and onground sections of Database Systems. This upper-division course provides an introduction to relational databases covering modeling, normalization, and design. As a control factor, all exams were taken in a proctored environment regardless of instruction method. Online content in the course was delivered through both e-College [2] and Blackboard [1] course sites and was readily available to both online and onground students. With over two years of data and 195 observations (134 onground / 63 online) we have applied two different statistical methods using the students’ final course grades as the measure of individual performance.

6.2 Methods of Analysis

In order to determine whether or not content delivery method significantly impacted student performance, we employ two techniques of analysis. The first of these is a simple comparison of means, a Student’s t-test. This test is used to determine if there is a significant difference between the means of two samples [4]. Simply put, it answers the question “Was there a difference?” Rather than looking solely at the mean course grade values for each sample, we first perform a number of t-tests on background characteristics, shown in table 1, for the two samples. This allows us to ensure that there were no pre-existing conditions that would otherwise account for variation between the samples.

Table 1: Demographic Variables

Variable	Description
ACTCOMP	Composite ACT Score
TOTHR	Number of total credit hours completed
GPA	Cumulative GPA
MJCAT1/2	Dummy variables used to describe a student’s major: CS, MIS, or Other

The second method used was regression analysis. This method attempts to model the input data linearly and determine which variables contribute significantly to the effectiveness of the overall model [4]. In general, it reveals which variables are important. Regression analysis also makes use of the demographic variables from Table 1, with the exception that ACT composite scores are dropped due to missing data values.

6.3 Results

To the demographic variables listed above, we add a variable for the student’s final letter grade in the course. In order to use the course grade for analysis, we map each letter to a numerical value where an A corresponds to 4 and an F is given the value 0.

Table 2 describes the makeup of the online and onground sections through the mean values of their demographics. Clearly, the two samples appear to have had very similar makeup. The average ACT and GPA scores are nearly identical.

Table 2: Mean Values for Demographics

Variable	Online		Onground	
	Mean	Std Dev	Mean	Std Dev
ACTCOMP	19.361	10.167	19.418	10.287
TOTHR	80.101	34.506	75.672	31.871
GPA	2.9913	0.5965	2.9919	0.7200

However, simply looking at the mean values on the surface is not enough to come to a conclusion. Table 3 illustrates the results of the various t-tests used during analysis. At the $\alpha = 0.01$ level, there are no significant differences between mean values for any of the background characteristics. This leads us to conclude that neither group had an unfair advantage in terms of the quality of students enrolled.

Table 3: Comparison of Mean Values

Variable	t-Value	P-value	Significant?
ACTCOMP	0.04	0.9712	N
TOTHR	0.88	0.3818	N
GPA	0.01	0.9958	N
GRADE	3.16	0.0019	Y

Given an equal footing on the demographic data, a legitimate comparison of the course grades can be made. With a confidence level of over 99%, we conclude that there was a significant difference between the mean final grades. The mean score for onground students was a 2.76 while the online students averaged a 2.24. Thus, onground students outperformed online students. Figure 1 depicts these two populations, with an assumed equal variance.

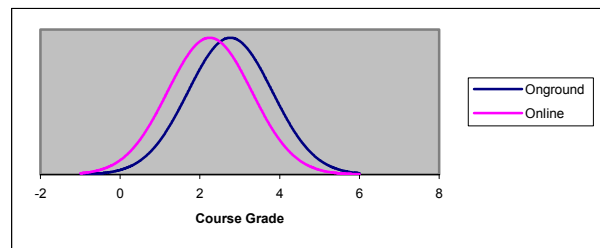


Figure 1: Online and Onground Populations

Regression results are presented in table 4. Each line depicts the best one through five variable models with the GRADE serving as the dependent variable. Values below each variable indicate the p-value for that variable entering the model last given the others.

Table 4: Regression Data

Variables/P-Values					R-Square	C(p)	
GPA					0.115	24.850	
<.0001							
GPA	ONLINE				0.164	14.893	
<.0001	0.0010						
GPA	ONLINE	TOTHRS			0.214	4.656	
<.0001	0.0003	0.0006					
GPA	ONLINE	TOTHRS	MJCAT2			0.220	5.210
<.0001	0.0003	0.0016	0.2308				
GPA	ONLINE	TOTHRS	MJCAT2	MJCAT1	0.225	6.000	
<.0001	0.0003	0.0021	0.1296	0.2728			

The various data items in Table 4 indicate a number of different characteristics about each model. P-values located below each variable indicate the degree of significance with which that variable adds to the model containing all other variables (within that row). Essentially it is used to determine if adding a variable is useful given the others. R-square (R^2) values generally are a sign of how well a linear model fits the sample data. These values are compared based on how close they are to the maximum R^2 value. Finally, the Mallows' Statistic or C(p) value is a measure of the bias for a model. If significant variables are missing from a model, the bias will be higher than ideal. We look for biases that are close to or less than the total number of parameters in the model.

Clearly, the single most important variable in the model is cumulative GPA. Not only does it make up the best one variable model, but with p-values less than .0001 when added last, it also is a significant contributor to all other models. The variable which denotes the course delivery method, ONLINE, entered in the two variable model adds to the predictive capacity of the model.

In general the three-variable model with GPA, ONLINE, and TOTHRS as the independent variables is the most attractive model for this data. P-values for each variable in the model indicate a high significance. It has a relatively large R-square value when compared to the maximum of 0.225. Lastly, it also has a C(p) value which is one of the first that is close to the number of parameters in the model at a given stage (4 in the case of a 3-variable model).

To summarize, t-tests show that onground database students significantly outperformed online students. Regression indicates that the variable for course type adds to the overall predictive capacity of the best model. Taken together, these results lead us to conclude that online and onground students did indeed perform differently, with onground students earning higher marks.

7. CONCLUSION

The results of our study clearly indicate that, for the student groups that we studied, online students are less successful than onground students. Beyond this basic fact, we must use caution in drawing any further conclusions. In the event that additional long-range studies yield similar results, it does not mean that online courses should not

be offered. Online learning skills are important for computer science students. If students do not learn as well online as onground, we need to determine the factors responsible for this difference and, based on that information, find ways to improve performance.

Computer science students are adept at using the web to find information. Online learning should be natural for them. However, it may be that our model of online learning is incorrect. The traditional model of face-to-face instruction is based on centuries of experience. The model the authors are currently using for online instruction is very similar to the onground model, absent the classroom. The same materials are used and the same types of assessments are used. It may be that we need to completely revise our model of online instruction.

Our next step will be to gather additional data to determine why performance is lower in the online courses and to try to modify the online model to compensate for those factors. In addition, we need to investigate new teaching models. Most importantly we need to remember that traditional classroom courses have a long history. It may take time to determine the appropriate online model.

8. REFERENCES

- [1] Blackboard website. <http://www.blackboard.com>
- [2] e-College website. <http://www.ecollege.com>
- [3] Hulme, M., and Locasto, M. Using the Web to Enhance and Transform Education, ACM Crossroads, 10.1, (Fall 2003), 23-28.
- [4] Kleinbaum, D. et al. Applied Regression Analysis and Other Multivariable Methods. PWS-Kent Publishing, Boston, 1988.
- [5] Kleinman, J., and Entin, E. Comparison of In-Class and Distance-Learning Students' Performance and Attitudes in an Introductory Computer Science Course. The Journal of Computing in Small Colleges: Proceedings of the Fifth Annual CCSC Northeastern Conference, 17,2 (April 2002), 206-219.
- [6] Korhonen, A., Malmi, L., Myllyselka, P., and Scheinin, P. Does it Make a Difference if Students Exercise on the Web or in the Classroom? SIGCSE Bulletin: Proceedings of the 7th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education, 34,3 (June, 2002), 121-124.
- [7] McCollum, K. A professor divides his class in two to test value of online instruction. Chronicle of Higher Education, 43, 23, 1997
- [8] McDonald, G., McDonald M., and Dorn, B. Teaching Without a Classroom: Delivering Courses Online in Proceedings of the Thirty-Sixth Annual Midwest Instruction and Computing Symposium (Duluth MN, April 2003), CD-ROM
- [9] MIT's OCW website. <http://ocw.mit.edu/index.html>
- [10] Olsen, D. A Comparison of Online and Lecture Methods for Delivering the CS 1 Course. The Journal of Computing Sciences in Colleges: Proceedings of the Fourth Annual CCSC Northwestern Conference, 18,2 (October 2002), 57-63.
- [11] University of Phoenix website. <http://onl.uophx.edu/programs/default.asp>