# Comparing Student Satisfaction and Perception of Effectiveness in Two Different Online Computer Science Courses

**Waleed Farag[1], Sanwar Ali[2], Imran Ghani[3]**

Indiana University of Pennsylvania, United States[1,2,3]

## Abstract

*This research attempts to find answers to the question: will changing course contents and difficulties have an impact on the level of students' satisfaction and perception of effectiveness in online courses? The paper focuses on presenting the detailed analyses and findings of indirect assessment techniques. Two courses (groups) are compared in this research: an introductory programming class versus a computer literacy one. The paper employs two different data sets and implements an experimental, in-depth analysis procedure to answer the stated research question. The first set uses data collected from students expressing their perception of the effectiveness of seven online course performance indicators. The second data set relies on data taken from a traditional student evaluation instrument to evaluate the level of students' satisfaction with the course and its instruction. The obtained results for most of the studied performance measures denote that there are no statistically significant differences between the two groups. However, the results also identify a few performance measures in which data in the two groups show statistically significant differences. Possible explanations of the obtained results are discussed. Lastly, brief results of direct assessment methods are also presented.*

**Keywords:** *Computer literacy, Online programming courses, Evaluating students' perception, Measuring students' satisfaction.*

## 1. INTRODUCTION

Online course delivery is a contemporary trend in education that has flourished over the past two decades. However, the exponential proliferation of such offerings poses many challenges [1] to overcome. Issues that need to be addressed include:

1. Measuring and improving the effectiveness of online courses.
2. Researching effective approaches to broaden applicability and improve retention rates.
3. Studying the ramifications of such wide-spread use on higher education.

The study presented in this paper is part of a multi-year, comprehensive research project that has the main objective of proposing a reliable framework for evaluating the effectiveness of online computer science course delivery. This proposed framework employs both direct and indirect assessment methodologies. In direct assessment techniques, data are derived based on student performance in class. Indirect assessment techniques use data collected from the students expressing their own perspective of course effectiveness. The analysis from applying indirect assessment techniques are discussed in detail in this paper, with results from direct assessment techniques briefly discussed in section 5.

Our initial research hypothesis can be stated as: Changing the course from an introductory programming to a computer literacy will have a significant effect on students' perception of effectiveness and overall satisfaction. This hypothesis was based on observations that some students find introductory programming classes more challenging, which will affect their satisfaction.

Several researchers have studied the evaluation of online courses. Helm et al. [4] introduced a systematic evaluation process. Rothman et al. [5] found that students rated content organization and format higher than other studied performance measures.

## 2. STUDY DESCRIPTION

This study compares two groups using an established framework for assessing the effectiveness of the courses previously described. The collected data for both groups were taken from courses that are offered fully online, taught by the same instructor, using the same Learning Management System.

The first group's data were taken from our Computer Literacy (CS 101) class. This course is an introduction to computers and covers basic topics such as computer hardware, operating systems, and MS-office applications.

The second group comprises data obtained from our introductory programming course (CS 110). This course is an introduction to programming, and covers basic programming concepts such as expressions, structured data types, and recursion concepts.

Although the contents in these two courses have different focuses, they have similar structures. Asynchronous activities in both courses include laboratory environments, online exercises, and discussion forums.

Our research hypothesis posed by this study is that the class change will result in some differences when measuring and comparing various course assessment factors.

## 3. ANALYSIS OF STUDENTS' PERCEPTION OF COURSE EFFECTIVENESS

Two surveys were designed to measure effectiveness. The first studies the students' approval of each of the numerous fundamental course components in effectively facilitating learning. The first set's questionnaires use a 4-level Likert-scale with values 1, 2, 3 and 4 coding the "Strongly Disagree", "Disagree", "Agree" and "Strongly Agree" responses, respectively. These fundamental course components/elements include the syllabus, projects, laboratory exercises and exams. The second survey proposed in [6] adopts a constructive learning model known as COLLES. This survey assesses the quality of the students' online experience in six central categories: relevance (relevant to professional practices), reflection (stimulation of critical thinking), interactivity (engagement in interactive dialogue), tutor-support, peer-support, and interpretation (making sense).

Questions in the later survey use a 5-level Likert-scale with values 1, 2, 3, 4 & 5 coding the "Almost Never", "Seldom", "Sometimes", "Often", and "Almost Always" responses, respectively. Under each of these six categories students rate four subcategories. These subcategory variables are combined to calculate a single figure that represents each category in the analysis (ex. relevance-110 is combined values of all 110 subcategories under relevance). Similarly, the eleven fundamental course elements of the first survey were combined into "elements". This reduction process yields seven distinct combined variables (see Table 1) that will be used to measure the students' perception of the efficacy of the course in all succeeding analyses.

We start the analysis by deriving simple descriptive statistics, then using the Shapiro-Wilk and Levene's tests to ensure that the data in both groups conform to parametric assumptions [3]. Normality test results in Table 1 show that the data for all factors are not normally distributed except for the interactivity factor, with p values for both 110 and 101 groups greater than .05, indicating normally distributed data. The Levene's test for the interactivity factor was non-significant denoting homogeneous variance. Thus, the parametric data assumptions hold for the interactivity factor. Mean values for the interactivity factor (see Table 1) might lead to an initial conclusion that the students in the 110 class have considerably positive perception of the degree of interactivity in the class, in comparison with students taking the 101 class. To investigate the statistical significance of such observation, the independent-samples t-test (Table 2) was used to compare both means for the interactivity factor. The t-test produced significant differences between the two studied groups, confirming the initial observation based on the mean values. The results of the t-test are formally reported below where SE is the standard error:

On average, students taking the 110 class ($\mu$ =3.26) have higher ratings of interactivity, than students taking the 101 class ($\mu$ =2.24), see Table 2. This difference was significant $t(68) = 3.23$, $p < 0.05$. Moreover, the effect size (r) is .37 which represents medium to large effect.

Table 1. Results for various statistical tests for both groups
Legend: μ= mean value, N=sample size, p=test significance, df=degree of freedom

| | μ | N | Shapiro-Wilk | | | Levene's test | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | test | df | p | test | df1 | df2 | p |
| elements–110 | 2.88 | 40 | .98 | 40 | .59 | 3.23 | 1 | 62 | .08 |
| elements –101 | 2.95 | 24 | .91 | 24 | .03 | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| relevance-110 | 3.57 | 46 | .94 | 46 | .03 | .10 | 1 | 68 | .75 |
| relevance-101 | 3.39 | 24 | .93 | 24 | .23 | | | | |
| reflect-110 | 3.85 | 46 | .94 | 46 | .02 | 1.02 | 1 | 68 | .32 |
| reflect-101 | 3.46 | 24 | .93 | 24 | .3 | | | | |
| interact-110 | 3.26 | 46 | .97 | 46 | .4 | .63 | 1 | 68 | .43 |
| interact-101 | 2.42 | 24 | .93 | 24 | .09 | | | | |
| tutor-110 | 3.07 | 46 | .85 | 46 | .00 | .46 | 1 | 68 | .50 |
| tutor-101 | 2.89 | 24 | .94 | 24 | .13 | | | | |
| peer-110 | 3.11 | 46 | .95 | 46 | .04 | 4.07 | 1 | 68 | .04 |
| peer-101 | 2.42 | 24 | .88 | 24 | .01 | | | | |
| sense-110 | 3.58 | 46 | .92 | 46 | .004 | 3.15 | 1 | 68 | .08 |
| sense-101 | 3.28 | 24 | .88 | 24 | .01 | | | | |

For the remaining six factors, Table 1 shows non-normally distributed data and/or non-homogeneous variances. Please also refer to Figure 1 which clearly depicts the non-normal distribution of the 110 data for the relevance factor showing negative skewness and kurtosis. To confirm whether the comparable mean values in Table 1 are statistically significant, the Mann-Whitney test was used (Table 2), and its formal result for relevance can be stated as:

> The students' perception of relevance in the 110 class (Median = 3.50) did not differ significantly from the same degree reported by students taking the 101 class (Median = 3.63), U = 504.00, z = -.59, ns (p = .55). The effect size (r) is -.07 which represents a small effect (see Table 2).

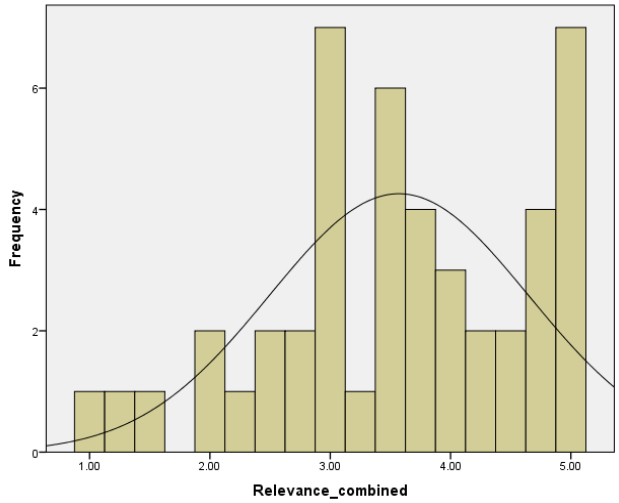


Figure 1. Non normal distribution of 110 relevance data
Legend: Frequency = frequency of ratings, Relevance_combined = combined values of relevance subcategories.

The Mann-Whitney test was also used to compare the mean values for the remaining five factors listed in Table 1. These tests produced insignificant results (see Table 2) except for the peer-support factor which can be reported as such:

> The students' perception of peer-support in the 110 class (Median = 3.13) differed significantly from the same rating reported by students taking the 101 class (Median = 2.75), U = 351.00, z = -2.51, p < .05. The effect size (r) is -.3 which represents a medium effect (see Table 2).

The results in Table 2 indicate that there were no significant differences between the students' perception of the remaining five performance indicator factors. Both the interactivity and peer-support factors have shown significant difference between the studied groups.

Student motivation and the course difficulties are two factors in explaining the obtained significant results. It is believed that the relatively difficult programming concepts in the 110 class encouraged students to be more involved in various discussion forums resulting in a higher and statistically significant difference for the peer-support factor. Note that analysis of withdrawal rate, one of the direct assessment factors discussed in section 5, denoted that 110 class exhibits statistically significant higher rates compared to the 101 class. This can further support the above interpretation by considering that students who continue in the 110 class are more likely to actively participate in and contribute to the class in order to enhance their chances of passing the course. Similar arguments may also apply for the explanation of the significant result obtained for the interactivity factor.

Table 2. Perception results for t-test and Mann-Whitney test
t-test Legend: μ=mean, SE=standard error
Mann-Whitney Legend: U=test value, z=z value, p=significance, r=effect

| **Perception factor** | **Descriptive values** | | **Test results** | | | |
|---|---|---|---|---|---|---|
| **t-test** | **μ** | **SE** | **t** | **df** | **p** | **r** |
| interact-110 | 3.26 | .14 | 3.23 | 68 | **.002** | .37 |
| interact-101 | 2.42 | .23 | | | | |
| **Mann-Whitney** | **Median** | | **U** | **z** | **p** | **r** |
| elements–110 | 2.89 | | 423.00 | -.79 | .43 | .10 |
| elements –101 | 3.17 | | | | | |
| relevance-110 | 3.50 | | 504 | -.60 | .55 | -.07 |
| relevance-101 | 3.63 | | | | | |
| reflect-110 | 3.88 | | 448 | -1.29 | .20 | -.16 |
| reflect-101 | 3.63 | | | | | |
| tutor-110 | 3.75 | | 509.50 | -.53 | .60 | -.06 |
| tutor-101 | 3.00 | | | | | |
| peer-110 | 3.13 | | 351.00 | -2.51 | **.01** | -.30 |
| peer-101 | 2.75 | | | | | |
| sense-110 | 3.75 | | 496.00 | -.70 | .48 | -.08 |
| sense-101 | 3.88 | | | | | |

## 4. EVALUATING STUDENTS' SATISFACTION IN BOTH GROUPS

The student evaluation instrument employed at our institution was used to perform the following analyses. For this study, eight of the 23 performance metrics in this instrument were selected, summarized as: course organized effectively, ideas communicated clearly, assistance provided, knowledge increased, feedback given, skills learned had value, student would pursue more classes taught by instructor, and overall instruction satisfaction.

The first seven metrics use a 4-level Likert-scale described in section 3. These seven metrics are combined to form one single factor denoted as course_satisfaction. The last metric uses a 5-level Likert-scale with values 5, 4, 3, 2, & 1 coding the "Superior", "Above Average", "Average", "Below Average", and "Poor" responses respectively, denoted as instruct_satisfaction.

We start by deriving basic descriptive statistics and test the conformance with the parametric data assumptions. The collected data are tested using Shapiro-Wilk normality and Levene's homogeneity of variance tests. The results of these tests yielded non-normally distributed data for both factors. Thus, the non-parametric Independent Samples Mann-Whitney test will be used for the analysis of both performance factors.

Table 3 lists descriptive statistics for the two factors in both groups and the Mann-Whitney test results. For course satisfaction and instructor satisfaction, there were no statistically significant differences (which opposes our initial research hypothesis).

Students' overall satisfaction levels with the instruction in the 110 class (Median = 4.00) did not differ significantly from the same levels reported by students taking the 101 class (Median = 4.00), U = 218.00, z = -1.69, ns (p = .09), see Table 3. The effect size (r) is -.24 which represents a relatively small effect.

Table 3. Mann-Whitney results for both satisfaction measures
Legend: μ=mean, SE=standard error, U=test value,
z=z value, p=significance, r=effect size

| Studied factor | Descriptive values | | Mann-Whitney results | | | |
|---|---|---|---|---|---|---|
| | μ | SE | U | z | p | r |
| 110 course_ satisfaction | 2.81 | .17 | 276.50 | -.47 | .64 | -.07 |
| 101 course_ satisfaction | 2.85 | .22 | | | | |
| 110 instruct_ satisfaction | 2.96 | .28 | 218.00 | -1.69 | .09 | -.24 |
| 101 instruct_ satisfaction | 3.63 | .21 | | | | |

## 5. DIRECT ASSESSMENT MEASURES

This section briefly comments on direct assessment metrics which compared the Intended Learning Outcomes (ILOs). The statistical analysis of the obtained data yielded comparable ILOs with no significant difference between the two groups. Moreover, analyses of a second set of performance measures focused on student success and interactivity also yielded comparable distributions in most metrics with two exceptions: the withdrew rate and content utilization in which significant association between the class type and each of these two metrics were spotted [2].

## 6. CONCLUSIONS

The paper presented a study (using indirect assessment measures) to determine whether course type and content difficulties have statistically significant effects on students' perception and satisfaction. First, the study assessed the perception of students via several specific performance metrics. Five of the analyzed seven measures showed non-significant differences across the distributions of the two studied groups while the interactivity and peer-support measures have exhibited significant differences. Such findings indicated that course type and difficulty might have a statistically significant influence on how students perceived certain performance measures. Second, we compared students' satisfaction levels when taking the 110 class with the same levels when taking the 101 class. The obtained statistical results indicated that there were no significant differences between the two groups for both adopted performance measures.

## REFERENCES

[1] Armitage, W., Boyer, N., Langevin. S., and Gaspar, A. "Rapid conversion of an IT degree program to online delivery: impact, problems, solutions and challenges.", SIGITE Conference on information technology education, New York, NY, ACM, 2009,100-107.

[2] Farag, W. and Ali, S. "Can Online Delivery Result in Comparable Achievement of Course Outcomes and Student Success in Different Computer Science Courses?", IEEE 2016 Frontiers in Education (FIE) International Conference, Erie, PA, 2016, 1-7.

[3] Field, A. "Discover Statistics Using SPSS", London, UK, SAGE Publications Ltd, 2017.

[4] Helm, J., Powell, K., and Ice, P. "Evaluating Online Course Quality for Student Learning and Success", World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, Honolulu, HI, 2011.

[5] Rothman, T., Romeo, L., Brennan, M., and Mitchell, D. "Criteria for Assessing Student Satisfaction with Online Courses" International Journal for e-Learning Security, 2011, 27-32.

[6] Taylor, P. and Maor, D. "Assessing the efficacy of online teaching with the Constructivist On-Line Learning Environment Survey", 9th Annual Teaching Learning Forum, Perth, Australia, 2000.