

A COMPARATIVE ANALYSIS OF ASSESSMENT RESULTS FROM FACE-TO-FACE AND ONLINE EXAMS

**Dr. Emiliya Koleva, Assist. Prof.,
Dr. Neli Baeva, Assist. Prof**

Nikola Vaptsarov Naval Academy – Varna (Bulgaria)

Abstract. In this study, a comparative analysis of the results of students' performance on a face-to-face and an online exam is made and presented. The students involved in the research are trained and evaluated by the same examiner. Different statistical tests are made using statistical analysis software. As a result of the research, the hypothesis is confirmed that there is a difference between the two evaluations. Comparison of the grades between the different exams showed that there is a linear relationship between them, there is dependence between the results from both exams and the results from the online exam are slightly higher than the results from the face-to-face exam.

Keywords: Statistical analysis; Students' performance; Exam evaluation comparison; Online exams; Face-to-face exams

1. Introduction

In recent years, the coronavirus disease pandemic 2019 (COVID-19) has changed the landscape of higher education. Due to this fact, in the last two years the learning process has been carried out online, hybrid or through blended teaching methods. The question arises about the results of this training. This need has led to a great deal of research in this area in recent years. According to the results of some of these studies, there is no significant difference in student performance in face-to-face and online exams (Larson & Sung 2009; Newlin, Lavooy & Wang 2005; Stack 2015). According to the others, the average assessment results increase in favor of online exams (Al Salmi, Al-Majeed & Karam 2019). Stevens et al.(2021) research shows that 37 from 91 studies (41%) found that online students perform better. Online education improved students' exam performance during the COVID-19 pandemic.

All the 69 participants in the study are students at Nikola Vaptsarov Naval Academy and they have held two exams: one online and one face-to-face. The course and the exams were conducted by the same professor. Both exams consisted of 4 free-response questions. The solutions were written by the students in their own handwriting. At the end of the online exam all the solutions were captured,

attached as files and sent to the examiner. During the online exam students were with switched on camera and microphone. They joined the exam from an account, given from the university. The handwriting from the two exams was compared, for each student. Both exams used the same grading scale.

In the present investigation, the research questions are:

1. Is there a linear relationship between the grades obtained in the two types of exams?
2. Is there a relationship between the two types of assessment?
3. Is there a statistically significant difference between the arithmetic mean of the results obtained from the online training and the results obtained from the face-to-face training and assessment?

The software used to make the statistical analysis in this study is SPSS. SPSS offers the ability to easily compile descriptive statistics, parametric and non-parametric analyses.

Method

A statistical hypothesis test is the method of statistical inference used in the study to decide whether the data at hand sufficiently support a particular hypothesis. The participants involved in the research are 69 second-year students. The study design consisted of an intragroup design – the same participants are measured more than once. Quantitative variables and interval scale were used.

Procedure

The data were taken from online students' assessment and face-to-face assessment in the form of written tests. The selected significance level is $\alpha = 0.05$. Descriptive statistics were made up by calculating the arithmetic mean of the two scales, the standard deviation, variance, mode and median. The linear dependency between the online and traditional rating scales is checked using the Pearson correlation coefficient.

In order to test the hypothesis of equality of the arithmetic mean between the grades received by the different exams, the following were performed:

1. Descriptive statistics;
2. χ^2 test;
3. Paired Samples Test.

2.Descriptive Statistics

Descriptive statistics describe, show, and summarize the basic features of a dataset found in a given study. It helps to understand the data better. The descriptive statistics used in this study such as the mean, standard deviation and variance, gives summary statistics about two continuous numeric variables – the assessment from face-to-face exam and online exam.

Table 1. Descriptive Statistics

	N	Mean	Median	Mode	Std. Deviation	Variance
Face-to-face	69	3,4638	3,0000	2,00	1,45096	2,105
Online	69	3,5362	3,0000	2,00	1,47109	2,164
Valid N (listwise)	69					

For both face-to-face exam grades and online exam grades, the mean, median and mode are all similar, indicating that the data are probably normally distributed (see Table 1).

The standard deviations indicate the extent to which the scores lie apart. For the spread of scores (as shown by the Std. Deviation) SPSS reports again similar results for face-to-face and online exam grades.

A descriptive statistic for the female ($n = 23$) and male students ($n = 46$) on exam grades is made. From the following two tables is obvious that the results received from the exams are similar for women and men (see Table 2 and Table 3).

Table 2

Report				
Face-to-face				
Gender	Mean	N	Std. Deviation	Variance
women	3,5217	23	1,64785	2,715
men	3,4348	46	1,36059	1,851
Total	3,4638	69	1,45096	2,105

Table 3

Report				
Online				
Gender	Mean	N	Std. Deviation	Variance
women	3,7826	23	1,70445	2,905
men	3,4130	46	1,34290	1,803
Total	3,5362	69	1,47109	2,164

The next step in the descriptive analysis is the construction and subsequent examination of a scatterplot.

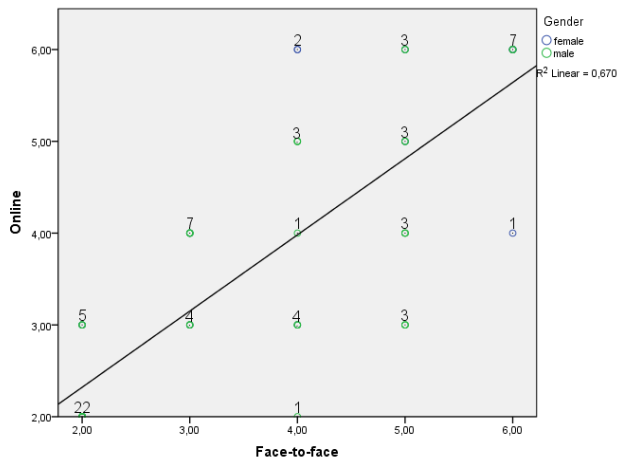


Figure 1. Scatter plot chart

The scatterplot on fig.1 suggests a definite relationship between the grades from Face-to-face and Online exams. There appears to be a positive correlation between the two variables. The R-squared value is **0.670** which means that the points are close to the linear trend line.

3. Correlation test

The Pearson correlation coefficient is a measure for the linear relationship between two quantitative variables. The research question is: “Whether or not a linear relationship between online and traditional assessment exists?” In this case the answer to the research question, concerning the existence of a linear relationship, is received by using the Pearson’s correlation test.

The first step is to specify the null and alternative hypotheses:

H_0 : there is no correlation between the student’s grades from face-to-face and online exams ($r = 0$).

H_1 : there is correlation between the grades ($r \neq 0$).

Table 4

Correlations			
		Face-to-face	Online
Face-to-face	Pearson Correlation	1	,819**
	Sig. (2-tailed)		,000
	N	69	69
Online	Pearson Correlation	,819**	1
	Sig. (2-tailed)	,000	
	N	69	69

** . Correlation is significant at the 0.01 level (2-tailed).

Results obtained by SPSS

A Pearson’s correlation was run to determine the relationship between 69 student’s grades from face-to-face and online exam.

SPSS reports (see Table 4) the p –value for this test being **0.000**. Because the p –value is smaller than the significance level $\alpha = 0.01$, the null hypothesis is rejected in favor of the alternative. So, the conclusion is that there is a very strong, positive correlation between the results from face-to-face and online exam ($r = 0.819$, $N = 69$, $p < 0.01$).

4. Chi-square test

The Pearson’s χ^2 test is the most commonly used test to find out whether the two categorical variables (in this case, Face-to-face exam and Online exam) are associated with each other – that is, are they dependent or independent? The chi

square test is appropriate for this task. The χ^2 test is used, instead of Fisher's exact test, when the sample size is bigger than 20.

The study presents two measures of students' test performance – Pass/Fail status, during face-to-face and online exam. The number of the student in the sample is 69. They took two exams during the semester – one at the university and one online.

The Case Processing Summary table (see Table 5) is a summary of the cases that were processed when the crosstabs analysis ran. There are 69 valid cases, and no missing cases.

Table 5. Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Face-to-face -exam * Online - exam	69	100,0%	0	0,0%	69	100,0%

The χ^2 test examining the relationship between the bad performance of students, during the face-to-face exam and the online exam and allows us to test this hypothesis. The test performs an independency test under following null and alternative hypotheses, H_0 and H_1 , respectively.

H_0 : These variables are not associated with each other – they are independent variables.

H_1 : The variables are associated with each other – they are dependent variables.

The first step of the chi-square test is the crosstab table. The crosstabs analysis is for the two categorical variables, Face-to-face exam, and Online exam. Each variable has two possible values: Fail and Pass for the Face-to-face variable; Fail and Pass for the Online exam variable. The received crosstabs table (see Table 6) includes information about observed counts and expected counts.

Table 6. Face-to-face -exam * Online exam Crosstabulation

			Online exam		Total
			fail	pass	
Face-to-face exam	fail	Count	22	5	27
		Expected Count	9,0	18,0	27,0
	pass	Count	1	41	42
		Expected Count	14,0	28,0	42,0
Total		Count	23	46	69
		Expected Count	23,0	46,0	69,0

The method of approximation used to calculate the chi-square test is reliable if the expected frequencies in cells are above 5 (see Table 6), or less than 20% of cells are above five (see Table 7). A good rule of thumb is that if the sample size ($n = 69$) is at least five times the number of cells (4) this should satisfy the final assumption.

Table 7. Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	46,274 ^a	1	,000	,000	,000
Continuity Correction ^b	42,783	1	,000		
Likelihood Ratio	52,513	1	,000	,000	,000
Fisher's Exact Test				,000	,000
N of Valid Cases	69				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 9,00.

b. Computed only for a 2×2 table.

The Table 6 shows that there is a big difference between the observed and expected counts. The question is whether these differences are big enough to conclude that the Face-to-face variable and Online variable are associated with each other. This is where the chi square statistic comes.

The chi square statistic appears in the Value column immediately to the right of "Pearson Chi-Square" (see Table 7). The obtained value is **46.274**.

The p -value (**.000**) appears in the same row in the "Asymptotic Significance (2-sided)" column. The result is significant if this value is equal to or less than the designated alpha level (normally **0.05**). In this case, the p -value is smaller than the standard alpha value, so the null hypothesis that asserts the two variables are independent of each other is rejected. So, the result is significant, which means that the variables Face-to-face exam and Online exam are associated with each other.

5. Dependent t-test

The dependent t -test is used to understand whether there was a difference between face-to-face and online assessment. The dependent variable is "exam grades", and the two related groups are the exam grades values received from "Face-to-face exam" and from "Online exam". The data must meet the requirements for a dependent t -test to give a valid result. In this case the normality assumption is not needed because the sample size is more than 30. During the check, five true outliers are received. They have a significant impact on the analysis, and they are removed from the data. So, the test is made for **64** valid cases (see Table 8).

Table 8. Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Online & Face-to-face	64	,893	,000

The dependent t –test compares the means between two related groups on the same continuous, dependent variable under following null and alternative hypotheses, H_0 and H_1 , respectively.

H_0 : The arithmetic mean values of the samples for the two estimation methods (Face-to-face and Online) are equal;

H_1 : The arithmetic mean values of the samples for the two estimation methods (Face-to-face and Online) are different.

The Paired Samples Test table (see Table 9) is where the results of the dependent t –test are presented. The information refers to the differences between the two exam grades.

SPSS reports the following results for the t –value(t), the degrees of freedom (df) and the significance level ($Sig. (2 - tailed)$): $p = 0,008 < 0,05$ - the H_0 hypothesis is rejected.

Due to the results, there was a statistically significant difference between Online and Face-to-face exam grades at $\alpha = 0.05$. This is because ‘Sig. (2-tailed)’ or $p < 0.05$. On average, online results were higher than Face-to-face results ($t = 2,741$).

Table 9. Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Online – Face- to-face	,23438	,68411	,08551	,06349	,40526	2,741	63	,008

Noorbehbahani et al. (2022) show in their research that one of the reasons for better student performance could relate to various cheating behaviors in online examinations. The study of (Watson & Sottile 2010) has reported that students are remarkably more likely to get answers from others during online exams or quizzes compared to live (face-to-face) ones.

Results obtained by Excel

Calculations were made and with Excel and the same results were obtained (see Table 10).

$$t(63) = 2,741$$

$$t_{critical} = t_{0,95}(63) = 1.99$$

$t > t_{critical}$ – the H_0 hypothesis is rejected.

Table 10

t-Test: Paired Two Sample for Means		
	Online	Face-to-face
Mean	3,578	3,344
Variance	2,280	2,039
Observations	64	64
Pearson Correlation	0,893	
Hypothesized Mean Difference	0	
df	63	
t Stat	2,741	
P(T<=t) one-tail	0,004	
t Critical one-tail	1,669	
P(T<=t) two-tail	0,008	
t Critical two-tail	1,998	

6. Conclusions

The study was focused on answering the research questions. In conclusion, the comparative analysis in the study shows that there is a very strong, positive correlation between the results from face-to-face and online exam. The result from the chi-square test shows statistically significant difference between the expected exam grades and the observed once. The *t* —test uses two highly correlated samples and report that there was a significant difference between their mean values in favor of online evaluation.

It was obvious that good students performed well either in paper exams or in online exams while poor students had difficulties in both types of exams. The performance of two exams is preserved but the grades are slightly inflated.

REFERENCES

- AL-SALMI, S. & AL-MAJEED, S. & AL-ZUBAIDY, S., 2019. Online Exams For Better Students' Performance. *9th International Conference on Education, Teaching & Learning (ICE 19)*, New York, USA.
- AL-QDAH, M. & ABABNEH, I., 2017. Comparing Online and Paper Exams: Performances and Perceptions of Saudi Students. *International Journal of Information and Education Technology*, **7**, 106 – 109.
- GANEVA, Z., 2016. *Da preotkriem statistikata s IBM SPSS Statistics*. Elestra, Available from: doi.org/10.13140/RG.2.1.2803.6080 [In Bulgarian].
- LARSON, D. & SUNG, C., 2009. Comparing student performance: Online versus blended versus face-to-face. *Journal of Asynchronous Learning*

- Networks*, **13**(1), 31 – 42. Available from: doi.org/10.24059/OLJ.V13I1.1675.
- NEWLIN, M. & LAVOY, M. & WANG, A., 2005. An experimental comparison of conventional and web-based instructional formats. *North American Journal of Psychology*, **7**(2), 327 – 336.
- NOORBEHBAHANI, F. & MOHAMMADI, A. & AMINAZADEH, M., 2022. A systematic review of research on cheating in online exams from 2010 to 2021. *Educ Inf Technol*. Available from: doi.org/10.1007/s10639-022-10927-7.
- PANDIS, N., 2016. The chi-square test, *American Journal of Orthodontics and Dentofacial Orthopedics, Statistics and research design*, **150**(5), 898 – 899, ISSN: 0889-5406, Available from: https://doi.org/10.1016/j.ajodo.2016.08.009.
- STACK, S., 2015. Learning outcomes in an online vs traditional course. *International Journal for the Scholarship of Teaching and Learning*, **9**(1), Available from: doi.org/10.20429/ijso.2015.090105.
- STEVENS, G. & BIENZ, T. & WALI, N. & CONDIE, J. & SCHISMENOS, S., 2021. Online university education is the new normal: but is face-to-face better?, *Interactive Technology and Smart Education*, **18**(3), Available from: doi.org/10.1108/ITSE-08-2020-0181.
- STOIMENOVA, E., 2000. *Izmervatelni kachestva na testove*, Institute of Mathematics and Informatics – BAS, Sofia, ISBN: 954-8986-07-8 [In Bulgarian].
- WATSON, G., & SOTTILE, J., 2010. Cheating in the Digital Age: Do Students Cheat More in Online Courses? *Online Journal of Distance Learning Administration*, **13**(1), ISSN: ISSN-1556-3847.

✉ **Dr. Emiliya Koleva, Assist. Prof.**

ORCID ID: 0000-0003-1902-2042

Web of Science Researcher ID: ACO-7965-2022

Faculty of Engineering

Nikola Vaptsarov Naval Academy

Varna, Bulgaria

E-mail: e.koleva@nvna.eu

✉ **Dr. Neli Baeva, Assist. Prof.**

ORCID ID: 0000-0002-3620-4692

Faculty of Engineering

Nikola Vaptsarov Naval Academy

Varna, Bulgaria

E-mail: n.baeva@nvna.eu