

Proctored vs Unproctored Online Exams in Language Courses: A Comparative Study

Mehmet Ali ÇELİKBAĞ* & Ömer DELİALİOĞLU

Department of Computer Education & Instructional Technology, Middle East Technical University, Turkey

*celikbag@gmail.com

Abstract: This study investigates online exams of online language courses at a distance education center of a public university in Turkey. Causal comparative research design was followed to examine issues in online language courses with the aim to better develop online exams. Participants were 105 students from spring semester of 2019 and 116 students from spring semester of 2020 from six different programs at associate degree level at a public university. The overall findings indicated that there were varied results in unproctored online exams (UOE) of online Turkish (written and verbal expression) and English (grammar) language courses. Effects of the COVID-19 pandemic were observed in an unproctored online final exam for the English language course. However, tests showed that there was a high reliability when they were administered in different years to similar conditions and groups of students. As an online exam platform, Moodle LMS was used. Students of Computer Programming and Justice programs in online language courses had greater achievement results when compared to other educational programs whether online or paper-based and proctored or unproctored exams were conducted. Although gender did not play a vital role in achievement results in general, it was highly influential in the unproctored online English final exam in favor of male students.

Keywords: Online exams, unproctored online exams, advantages and disadvantages of online exams, online language learning, distance education, COVID-19 pandemic effects on education

1. Introduction

The proportion of distance education is increasing, and this creates implications for design and implementation of courses by considering new challenges and advantages. Recently, the COVID-19 pandemic has created further implications in distance education, especially in terms of examination processes. This study aims to investigate online exams with the aim of finding ways to improve assessment experience and quality, and to analyze the effects of settings and conditions. Within the scope of this study, quality, reliability and validity issues regarding online summative assessment in distance education are discussed in a holistic way. Specifically, the study seeks to provide insights for effective design of online exams. Distance education and its evaluation for the improvement online exams are investigated with causal-comparative study. Since the beginning of the pandemic many countries have switched to open and distance education settings. In this study, six educational programs at a distance education center of a public university in Turkey were studied.

2. Literature Review

Online language learning is increasingly being welcomed in formal and informal education environments. Since Open University's online French course in 1995 and the last decade of 20th century, many developments have taken place such as learning languages online through MOOCs, mobile applications in online learning, especially in informal settings, and virtual worlds (Hockly, 2015). In the context of Turkey, online exams became prevalent since 2015 in distance education centers. Ilgaz and Afacan Adanir (2019) performed an analysis of online exams comparing them with traditional exams in literacy, foreign language and history courses in addition to surveying perception

of students about online exams. Their results showed that there were statistical differences between online midterm exams and traditional final exams of said courses. Hollister and Berenson (2009) conducted research to find differences between proctored and unproctored test performances of groups of students and found no difference regarding performance. Rios and Liu (2017) have also indicated the extensive use of unproctored exams in online education due to financial and flexibility issues while focusing on necessity of online proctoring facilities. In this sense, it might be important to balance available resources and desired outcomes. Recently, unproctored and proctored online exams have also gained popularity due to the COVID-19 pandemic.

2.1 Advantages and Disadvantages of Unproctored Online Exams (UOEs)

Online exams, specifically UOEs create new advantages that are not available in traditional exams. First of all, grading process can be automatically completed in online exams. In this sense, students are able to access their exam scores immediately, and receive feedback if given. Online exams can be reutilized with just a few clicks and eliminate all processes of exam preparation and printing that is carried out in traditional exams; therefore, it is indeed cost and time efficient. UOEs hold a crucial role in which students can access exam platforms with ease through low level requirements. Overall, UOEs not only provide advantages for learners but also for educators.

Although there are some major advantages in online exams of language courses, there are naturally some disadvantages. First of all, it may be difficult for instructors to set up question pools to be used in online exams as well as training instructors and learners for online exams (Clark et al, 2020). In transition from face-to-face traditional assessment to online assessment, academics need to review their assessment considerations and techniques (Hollister & Berenson, 2009). All this might be a time-consuming process for instructors and may require additional support. A contemporary server with adequate network bandwidth, CPUs and physical memory may be required. All these processes may increase initial costs of setting up an online exam environment. Additional disadvantages of these exams could be added stress on some learners who do not feel confident with technology. Goertler and Gacs (2018) similarly assert being successful in online language learning also depends on being competent in technology to some extent. This may affect overall assessment process in terms of validity.

Reliability and validity are the first issues to be considered in exams. Dermo (2009) indicates when an assessment is marked by computers, reliability of a test may increase which can be considered a further advantage of online assessment. Similarly, the reliability of a test results from producing statistically consistent measures when sampling error is eliminated (Dennick, Wilkinson, & Purcell, 2009). Therefore, the means and standard deviations of the grades of students registered at the distance education center subject to this study should be within acceptable boundaries affecting 20% of overall grade of students. UOEs with randomized questions require intense initial effort in order to maintain reliability and validity. Nonetheless, it may leave validity issues unanswered to some extent even though a qualified committee prepare question banks. In this sense, validity is a critical issue in all kinds of exams whether paper-based or online.

3. Method

3.1 Research Design

It is not always possible to manipulate independent variables in studies, yet natural conditions may take place such as a pandemic affecting dependent variables (Schenker & Rumrill, 2004). That is, causal-comparative research tries to “find relationships between independent and dependent variables” resulting from an action which is not possible to occur in normal conditions due to ethical considerations and regulations (Salkind, 2010, p. 124). In usual settings, UOEs are commonly used at distance education centers in Turkey; however, unproctored online final exams (UOFEs) were unexpected which were caused by the pandemic. Nonetheless, it may have less influence on distance education programs in comparison to traditional programs on campuses that have undergone a transition due to the COVID-19 pandemic. Since the following groups as categorical variables are

easily available, a causal-comparative design is adopted in overall design of this research: different educational programs, gender, academic year, paper-based proctored final exams, unproctored online midterm exams (UOMEs), and UOFEs. The research questions that shape this study are presented below:

RQ1: Do students in online language courses have higher achievement UOMEs when compared to proctored paper-based finals exams?

RQ2: Do students in online language courses have higher achievement in UOMEs when compared to UOFEs?

RQ3: Do students in online language courses have higher achievement in proctored paper-based final exams of 2019 prior to the COVID-19 pandemic when compared to UOFEs of 2020 during early stages of the COVID-19 pandemic?

RQ4: Do students in online language courses have statistically different achievement scores in UOMEs when compared to proctored paper-based final exams across different educational programs?

RQ5: Do students in online language courses have statistically different achievement scores in UOMEs when compared to proctored paper-based final exams across gender?

3.2 Participants

The participants of the study were students of online programs at a public university in Turkey. The data collected through convenience sampling has been given below:

Table 1. *Demographic Information of Participants of 2019 and 2020 Spring Semesters*

Year	Variables	Frequency		Percentage (%)	
		2019	2020	2019	2020
Gender	Female	70	72	66.7	62.1
	Male	35	44	33.3	37.9
Educational Program	JSTC	8	15	7.6	12.9
	BAI	9	20	8.6	17.2
	CP	21	27	20.0	23.3
	LOMS	14	13	13.3	11.2
	MDS	46	37	43.8	31.9
	THBA	7	4	6.7	3.4
Total		105	116	100	100

Notes: JSTC: Justice; BAI: Banking and Insurance; CP: Computer Programming; LOMS: Law Office Management and Secretarial; MDS: Medical Documentation and Secretarial; THBA: Tourism and Hotel Business Administration.

3.3 Procedures

Twenty multiple-choice questions were selected randomly from question pools and appeared on the screens of learners in two online language courses at college level. In spring semester of 2019, final exams were taken in the form of paper booklets and exam results were obtained from optical answer sheets. Data was gathered from Moodle database in terms of exam score, gender and educational program. Once data was consolidated in spreadsheets for SPSS analyses, all details of student numbers and names were removed. Ethical procedures were followed by acquiring permission from the institutional review board of the Middle East Technical University, and the participants' rights and confidentiality were protected.

3.4 Data Analysis

Data were analyzed by using SPSS and utilizing nonparametric tests, as data did not adhere to normal distribution, and transforming data did not lead to normal distribution. Wilcoxon Signed-Ranks Tests, Mann-Whitney U Tests, and Kruskal-Wallis Tests were employed to answer corresponding research questions.

4. Results

The quantitative data from Moodle LMS exam logs were analyzed through descriptive statistics and presented by providing mean, standard deviation, and frequencies in the form of tables. Means and standard deviations of two courses of 2019 and 2020 spring semesters can be found in two tables below:

Table 2. *Comparison of Language Exams for Six Educational Programs of 2019*

Educational Program	Turkish Language Midterm Exam (Online & Unproctored)		Turkish Language Final Exam (Paper-based & Proctored)		English Language Midterm Exam (Online & Unproctored)		English Language Final Exam (Paper-based & Proctored)	
	M	SD	M	SD	M	SD	M	SD
JSTC	76.87	7.98	81.25	9.54	82.50	28.41	69.37	26.38
BAI	66.66	13.69	66.11	14.52	63.33	16.39	27.77	21.37
CP	72.85	21.71	76.66	13.07	83.80	14.73	66.90	21.18
LOMS	70.35	19.46	65.71	11.57	61.42	18.75	29.28	15.42
MDS	72.93	14.51	71.08	11.10	72.71	19.93	47.82	18.96
THBA	54.28	13.97	53.57	15.19	67.14	24.64	42.85	20.17

Table 3. *Comparison of Language Exams for Six Educational Programs of 2020*

Educational Program	Turkish Language Midterm Exam (Online & Unproctored)		Turkish Language Final Exam (Online & Unproctored)		English Language Midterm Exam (Online & Unproctored)		English Language Final Exam (Online & Unproctored)	
	M	SD	M	SD	M	SD	M	SD
JSTC	78.00	10.48	74.00	9.67	73.33	19.88	73.00	22.02
BAI	63.25	15.83	63.25	12.69	69.50	29.15	67.50	24.52
CP	77.96	12.34	77.22	12.03	82.96	17.05	84.81	15.09
LOMS	62.30	16.28	68.46	11.43	53.84	22.92	57.69	25.46
MDS	70.54	12.29	67.83	13.92	70.13	18.04	69.18	21.13
THBA	62.50	17.07	57.50	13.22	88.75	14.36	70.00	19.57

4.1 Achievement Across Unproctored Online Midterm (UOME) and Proctored Final Exams

To evaluate statistical difference between UOME and proctored paper-based final exam scores of Turkish language course in the spring semester of 2019 a Wilcoxon Signed-Ranks Test was carried out which indicated that there were no statistically significant differences between UOME scores and paper-based proctored final exam scores $T = 2031$, $z = -.247$, $p = .805$. Similarly, for the English language course in the spring semester of 2019 a Wilcoxon Signed-Ranks Test indicated that there were statistically significant differences between UOME scores and paper-based proctored final exam scores $T = 300$, $z = -7.485$, $p < .001$. That is, students performed worse in proctored paper-based final exam compared to UOE of English language course in 2019 before COVID-19 pandemic.

4.2 Achievement Across Unproctored Online Midterm (UOME) and Unproctored Online Final Exams (UOFE)

To evaluate any statistical difference between UOME and UOFE scores of Turkish language course in the spring semester of 2020 a Wilcoxon Signed-Ranks Test was run and showed that there was no statistically significant difference between UOME and UOFE scores $T = 2444$, $z = -.446$, $p = .656$. Similarly for the English language course in the spring semester of 2020 the same test showed there were no statistically significant differences between UOE midterm scores and UOE final scores $T = 2437$, $z = -.134$, $p = .894$. That is, students performed similarly in both cases of UOE instances at different times during the pandemic.

4.3 Achievement Comparison Prior and During the COVID-19 Pandemic

For exam scores for the years 2019 and 2020, Mann–Whitney U tests indicated that the achievement scores only statistically differed between proctored paper-based English language and UOFE, $U(N_{2019} = 105, N_{2020} = 116) = 44.18, p < .001$. The median of the English language final exam achievement scores in 2020 ($Md = 80$) was higher than the median in 2019 ($Md = 50$).

4.4 Achievement Across Different Educational Programs

For the UOME for the year 2019, a Kruskal-Wallis test revealed that there was no statistically significant difference in the UOME scores of the Turkish language course across six educational programs, $H(5) = 10.57, p = .061$. Medians of exam scores across different educational programs were, to some extent, similar and were JSTC ($Md = 80$), BAI ($Md = 70$), CP ($Md = 80$), LOMS ($Md = 68$), MDS ($Md = 75$), THBA ($Md = 55$). However, when it was the proctored paper-based final exam for the year 2019, a Kruskal-Wallis test revealed a statistically significant difference in proctored paper-based final exam scores of the Turkish language course across six educational programs, $H(5) = 21.55, p = .001$. Two pairs of educational programs according to Dunn’s pairwise tests statistically differed. JSTC-THBA and CP-THBA were these pairs, and there was strong evidence ($p = .005$, adjusted using the Bonferroni correction) between JSTC and THBA, and ($p = .002$) between CP and THBA. JSTC ($Md = 82$) and CP ($Md = 80$) performed better than THBA ($Md = 60$).

For the UOME for the year 2020, a Kruskal-Wallis test revealed a statistically significant difference in UOME scores of the Turkish language course across six educational programs, $H(5) = 21.16, p = .001$. Dunn’s pairwise tests indicated two pairs of educational programs statistically differed which are CP-LOMS and CP-BAI. There was some evidence ($p < .05$) for both between CP-LOMS and CP-BAI, and CP ($Md = 85$) performed better than BAI ($Md = 65$) and LOMS ($Md = 60$). Similarly, for the UOFE for the year 2020, the same test revealed a significant difference in UOFE scores of the Turkish language course across six educational programs, $H(5) = 18.55, p = .002$. Only one pair of educational programs statistically differed, and it took place between CP and BAI. There was somewhat strong evidence ($p < .01$) between CP and BAI, and CP ($Md = 80$) performed better than BAI ($Md = 62$).

For the UOME for the year 2019, a Kruskal-Wallis test revealed a statistically significant difference in UOME scores of the English language course across six educational programs, $H(5) = 16.53, p = .005$. Only one pair of educational programs statistically differed, CP and LOMS. There was some evidence ($p = .020$) between CP and LOMS, and CP ($Md = 85$) performed better than LOMS ($Md = 65$). Similarly, for the proctored paper-based final exam for the year 2019, the same test indicated a significant difference in proctored paper-based final exam scores of the English language course across six educational programs, $H(5) = 34.15, p < .001$. Four pairs of educational programs differed, CP-BAI, JSTC-BAI, CP-LOMS and JSTC-LOMS. There was strong evidence ($p < .005$) between these pairs, and CP ($Md = 70$) and JSTC ($Md = 78$) performed better than BAI ($Md = 20$) and LOMS ($Md = 25$).

For the UOME for the year 2020, a Kruskal-Wallis test revealed a statistically significant difference in UOME scores of the English language course across six educational programs, $H(5) = 18.01, p = .003$. Only one pair of educational programs differed, CP and LOMS. There was strong evidence ($p < .005$) between CP and LOMS, and CP ($Md = 80$) performed better than LOMS ($Md = 70$). Similarly, for the UOFE for the year 2020, the same test indicated a statistically significant difference in proctored paper-based final exam scores of the English language course across six educational programs, $H(5) = 16.25, p = .006$. Two pairs of educational programs statistically differed which are CP-LOMS and CP-MDS. There was some evidence ($p < .05$) for both between CP-LOMS and CP-MDS, and CP ($Md = 90$) performed better than LOMS ($Md = 45$) and MDS ($Md = 70$).

4.5 Achievement Across Gender

Mann–Whitney U tests for gender influences on each of all four exams of English language courses showed no statistically significant differences except in one of the UOFEs in 2020. At this time, the Mann-Whitney U test revealed that UOFE scores of male students ($Md = 88, n = 44$) in English language course were higher than ones of female students ($Md = 70, n = 72$), $U = 1118, z = -2.66, p = .008$.

5. Discussion and Conclusion

The literature review shows that online language learning and online assessment, specifically UOEs, provide challenges and opportunities for both learners, instructors and institutions. This poses many comparably new phenomena to consider especially for stakeholders who do not have experience in online learning and teaching. Creating reliable and valid question categories, preventing cheating, and easing examination process for all are just some of concerns. As this study focused on UOEs in online language courses, it revealed some varying results depending on different research questions. In only one case, students had higher achievement in UOMEs when compared with proctored final exams, and it took place between UOME and proctored paper-based final exam of online English language course in 2019. Although this could be regarded as a result of cheating, it cannot be certainly known and asserted due to the limitations of this study. When both midterm and final exams were held in unproctored and online environments, there were no statistical differences between midterm and final exams of each language course. When two academic semesters prior and during COVID-19 pandemic were compared, it was understood that students had higher achievement scores in unproctored online English language final exam during the pandemic. Educational programs of CP and JSTC continuously showed higher achievement results in different situations. Gender had an impact on achievement of unproctored English language exams during early period of COVID-19 and the situation was in favor of male students. In the case of UOEs, it may be further necessary to detect gender influences in different domains in order to address underlying issues. That is, online learning environments and online learner populations continue to increase, research studies are to be conducted in different domains and by taking into account different considerations. Further research interests related to online exams may focus on carrying out investigations on enabling safe exam browsers, disabling navigation of questions, disabling text-copy feature, and utilizing proctoring software during exams such as screen, voice, and camera recording, all of which were not utilized in UOEs of this study. Regardless, there appears to be many advantages of UOEs indicated across the literature. Therefore, it is highly likely that UOEs will be utilized in the future, especially for online midterm exams in distance education programs with cost-efficient examination designs in which overall achievement of a student taking a course does not solely depend on passing an online exam.

References

- Clark, T. M., Callam, C. S., Paul, N. M., Stoltzfus, M. W., & Turner, D. (2020). Testing in the Time of COVID-19: A Sudden Transition to Unproctored Online Exams. *Journal of Chemical Education*, 97(9), 3413–3417.
- Dennick, R., Wilkinson, S., & Purcell, N. (2009). Online eAssessment: AMEE Guide No. 39. *Medical Teacher*, 31(3), 192–206.
- Dermo, J. (2009). E-Assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology*, 40(2), 203-214.
- Goertler, S., & Gacs, A. (2018). Assessment in online German: Assessment methods and results. *Die Unterrichtspraxis/Teaching German*, 51(2), 156-174.
- Hockly, N. (2015). Developments in online language learning. *ELT Journal*, 69(3), 308-313.
- Hollister, K. K., & Berenson, M. L. (2009). Proctored versus unproctored online exams: Studying the impact of exam environment on student performance. *Decision Sciences Journal of Innovative Education*, 7(1), 271-294.
- Ilgaz, H., & Afacan Adanır, G. (2019). Providing online exams for online learners: Does it really matter for them? *Education and Information Technologies*, 25(2), 1255-1269.
- Rios, J. A., & Liu, O. L. (2017). Online proctored versus unproctored low-stakes internet test administration: Is there differential test-taking behavior and performance? *American Journal of Distance Education*, 1-14.
- Salkind, N. J. (2010). *Encyclopedia of research design*. Los Angeles, CA: SAGE.
- Schenker, J. D., & Rumrill, P. D. (2004). Causal-comparative research designs. *Journal of Vocational Rehabilitation*, 21(3), 117-121.