

Online Assessment in Language Teaching Environment through Essays, Oral Discussion, and Multiple-Choice Questions

Liqaa Habeb Al-Obaydi (liqaa.en.hum@uodiyala.edu.iq) *Corresponding Author
English Department, College of Education for Human Sciences,
University of Diyala, Iraq
<https://orcid.org/0000-0003-3991-6035>

Marcel Pikhart (Marcel.pikhart@uhk.cz)
Department of Applied Linguistics, Faculty of Informatics and Management
University of Hradec Kralove, Czech Republic
<https://orcid.org/0000-0002-5633-9332>

Ragad M Tawafak (raghad@buc.edu.om)
Department of Information Technology, AlBuraimi University College, Oman
<https://orcid.org/0000-0001-8969-1642>

Abstract

Assessment has always played an essential role in any kind of education, including foreign language teaching, as it provides feedback to the students, enables reporting, and ensures progress development. Still, it can also be the cause of various forms of anxiety and stress. In the era of digital learning when basically all teaching processes can be conducted online, assessment can prove to be a major challenge for instructors. The research was conducted with two groups of students to evaluate whether multiple-choice questions or essay writing questions create any major advantage or disadvantage for the student's cognitive development and learning strategies. The results clearly show that there are differences between the two groups. The experimental group that was tested by using essays for their evaluation reached better results in their final oral exam, while the control group that was tested by only using multiple-choice tests reached better results in their interim evaluation, but they showed only lower levels of cognitive skills. The results show that to improve cognitive and oral competencies, it is more efficient to assess the students with written essays via weekly quizzes in addition to oral discussion tests rather than just fill-in-the-form tests despite the fact that these tests are able to generate better results in the students' formative exams.

Key words: Language assessment, Essays questions, Multiple-choice questions, EFL environment, Online education

Introduction

The transition of various courses, including foreign language learning (FLL), to the online environment started a few years ago and has recently been further expedited by the COVID-19 pandemic. Stakeholders and teachers were compelled to convert to online teaching immediately as a substitute option and this reality spread around the

globe and radically changed the way teachers teach and students learn (Çakmak et al., 2021; Turnbull et al., 2021). Consequently, the COVID-19 lockdowns also had a significant impact on the English as foreign language (EFL) learners' instruction and their further assessment. An assessment can be understood as “the heart of the student experience”, and also “the single biggest influence on how students approach their learning” (Rust et al., 2005, p. 231). Therefore, this study tries to shed light on online assessment in more details so that it can be optimized for online FLL classes.

Formative and summative assessments are the two types of evaluation that are most frequently utilized in FLL. Formative assessment is the type of evaluation that emphasizes using information and feedback to enhance learning process (Arend, 2006; Kremmel & Harding, 2019). Summative assessment, on the other hand, refers to a formal evaluation of learning attainment that is carried out after the courses, lessons, or projects finish (Sewell, 2010). Naturally, both can be used to evaluate students' learning progress but they are used with different intentions. Previously, those two assessment possibilities were utilized in the FLL classrooms, but in recent years, they have also been used for online FLL due to the fact that all, or almost all, FLL is conducted online.

Teachers have more opportunities to do online evaluation in the era of digital media and it is almost obligatory to use various modalities of online education in their everyday teaching practice. The advantages of online assessment, for instance, include the ability to evaluate crucial life skills, increase the validity of the results and, as a result, to caliber the test itself, and help teachers avoid the disadvantages of the conventional paper-based assessment system, such as the time needed for grading (Hamadah, 2017; Zhang et al., 2021). According to Seifert and Feliks (2018) online tests also reduce the need for paper and the time needed to print assignments. The benefits of online assessment include accessibility from anywhere at any time on any device, reliable and quick feedback, automatic grading and reporting. In addition to providing more tools than in a regular classroom to make the process easier, this approach gets beyond traditional learning time and space constraints. It enables students to study independently, releasing them from the constraints of conventional educational methods (Wang, 2014). Additionally, it encourages students to participate, it provides feedback to more students, saves time, provides high-quality test data, lowers the cost of printing, and improves objectivity in grading because the computer grades exams without taking into account the names, races, or cultures of the students (Hamadah, 2017; Yulianto & Mujtahin, 2021). Online assessment also offers rapid feedback, and the process of scoring immediately reduces the time of teachers required for manual data entry. Multiple choice questions (MCQ), group projects, online discussions, essay questions, team case studies, and self-evaluation are core modalities of online assessment (Khairil & Mokshein, 2018; Zhang et al., 2021).

Though there are many advantages of online assessment, there are also several disadvantages. Many researchers concluded that teachers can face many challenges in dealing with assessment online. Abduh (2021), Hricko & Howell (2006), and Yulianto and Mujtahin (2021) reported that online tests put more demand on a variety of abilities, including typing, using numerous screens to recall a passage, mouse navigation, and the usage of combinations. They add that reading text on a computer screen might sometimes make some individuals more tired than reading information on paper, long passages on a computer screen could be more challenging to read, it is difficult to see a

whole page or document on the screen at once since some elements require horizontal and vertical scrolling to display an entire graphic, students with visual impairments also face significant challenges while using graphic user interface. But it is also true that various conversion tools and other web tools are continually being created and improved but some of these features might not be available to everyone. Arif (2020) adds that teachers encountered a number of difficulties, including trouble scoring, internet connectivity, and misinterpretation of the supplied instruction. He revealed that the teachers came up with some special ideas and personal solutions to the issues they encountered.

With the sudden obligation of online education and online assessment as a result of global pandemic of COVID-19, many institutions around the world were not prepared or equipped enough to deal with online assessment professionally (Pikhart & Al-Obaydi, 2023). So, they depended on the simple kinds of assessments (traditional methods) to evaluate the academic performance of the students. Some of these kinds are MCQs. Though they were considered as the most predominant kind of online tests, they are not yet sufficiently examined (Kim, Smith & Maeng, 2008). Many researchers around the world used MCQs when measuring only lower levels of cognitive skills and the impact of this sort of online evaluation, which is the simplest to use, on the stimulation of study strategies and the accomplishment of necessary learning objectives is yet not fully understood (Aryadoust et al., 2021; Nowicki & Jones, 2005; Oliver & Dobeles, 2007; Scouller, 1998; Shumway & Harden, 2003).

Additionally, it is debatable if alternative online knowledge evaluation techniques, essay questions for instance, support the development of deep learning strategies and the accomplishment of more advanced learning objectives and help for more engaging activities, as supported by Al-Obaydi et al. (2023) who claimed that involvement in online discussions, written or oral in activities such as structured feedback, could have acted as a catalyst to encourage pupils to learn more deeply by gaining different types of engagement. In addition, to create a successful online assessment, teachers must increase and enlarge the measurements used throughout the online course's instructional delivery (Robles & Brathen, 2002; Yulianto & Mujtahin, 2021).

Based on that, this study attempts to offer more insights regarding the efficacy of online assessment. The main idea behind the present experiment was to evaluate the differences in EFL assessment when conducted online through two possible modalities, i.e., written essays with oral discussion and multiple-choice questions (MCQ). Two groups, experimental and control were formed to evaluate the differences in their performance, level of understanding, speaking and cognitive skills, and their overall language performance when using essays or MCQ for their monthly assessments. It is essential to mention that only a limited amount of research has been done, leaving room for future studies to confirm or disprove the accuracy of some traditional (face-to-face) teaching and assessment-based concepts in the context of the new online education context. Thus, using essay questions with oral discussion as an alternative to MCQs, as suggested by the present study, seems worth analyzing.

To understand the situation and verify if MCQ will have different results compared to essay and oral discussion questions as a test method, the following hypotheses were formulated and then verified by the subsequent research.

1. There is no difference between the experimental group and the control group in the results of the first monthly exam.
2. There is no difference between the experimental group and the control group in the results of the second monthly exam.
3. There is no difference between the experimental group and the control group in the results of the third monthly exam.
4. There is no difference between the experimental group and the control group in the results of the oral discussion exam.
5. There is no difference between the experimental group and the control group in the development of cognitive skills as a result of the kind of testing used.

Literature Review

The two evaluation kinds that are most frequently used in FLL are summative and formative assessments. Formative assessment is a style of evaluation that places an emphasis on using data and feedback to improve the learning process. On the other hand, summative assessment describes a formal assessment of learning achievement that takes place after the courses, lessons, or projects are completed. Naturally, both can be used to gauge pupils' progress in their learning, but their purposes differ. Language assessment, whether summative or formative (paper-based or online), has recently received much scholarly attention (Hildén & Fröjdendahl, 2018; Kremmel et al., 2018; Kremmel & Harding, 2019). This is because any kind of assessment is crucial for optimal learning process and its outcomes. It is crucial for raising the standards of instruction and improving the learning process. Additionally, it also encourages students to learn, provides feedback on their development, assesses their success, and pinpoints the necessary curriculum advancement. Due to the special circumstances that faced the educational process during the pandemic of COVID-19, the use of online assessment becomes natural among teachers and students. The use of mobile devices and social media, interaction with virtual characters like avatars, sophisticated online training for raters of writing or speaking tests, and automated scoring of extended writing are all features of online testing that offer more or less instantaneous results and diagnostic feedback. There are numerous types of online tests described, including multiple-choice, fill-in-the-blank, short answer, performance, conferencing, portfolio, and self-assessment tests (Maley & Kiss, 2018). Most of these types were used as the determined need of the students and depending on other factors as reported by Zhang et al. (2021) who examined the methods used by six EFL instructors at a Chinese university for online evaluation. The study's conclusions showed that the EFL teachers chose particular evaluation techniques and made assessment decisions based on local policy, their own teaching experience, and subjective decisions.

Sutadji et al. (2021) utilized a variety of assessment methods, including a written test with case study questions and analysis, online discussions with peer evaluation, and teacher evaluation of student activity. The study's conclusions showed that using

various evaluation methodologies led to the creation of a more authentic assessment. Marzano and Kendall (2007) presented one of the most thorough overviews of how well a specific sort of knowledge evaluation may be used to gauge the success of a specific level of educational goals. Their findings suggest that a relatively small number of educational goals can be adequately assessed using multiple-choice questions, whereas a far larger range of educational goals, all higher-level goals and the majority of lower level goals, can be assessed using essay type questions. Zlatović et al. (2015) extended the ideas when led a research by testing and surveys of 351 students from higher education institutions. The findings show that by utilizing different types of online assessments, students' learning strategies can be influenced relatively quickly, and that encouraging deeper learning strategies has a positive effect on the formal and perceived success rates in accomplishing the intended learning objectives. They concluded that a surface learning approach is stimulated by the announcement and use of online tests with multiple-choice questions, and a deep learning strategy is only slightly less stimulated.

The introduction and use of online assessments in the form of essays encourages the use of deep learning strategies, and to a lesser extent, surface learning strategies. The difficulties and potential solutions related with converting a group speaking assessment task from face-to-face to online mode after the pandemic's emergence at a university setting in Hong Kong were researched by Forrester (2020). The outcomes revealed that teachers thought the new one-to-one conversation assessment was good. The pupils were divided, with some favoring of the first format of the group discussion. Yulianto and Mujtahin (2021) studied the opinions and practices of the ELT teachers related to the use of online assessment during COVID-19. The authors discovered that the teachers had negative feelings about the online evaluation during the COVID-19 epidemic through open-ended surveys and online interviews. They were especially worried about the pupils' lack of excitement, the veracity of their assessment, and the internet connection. On another side, Shen, Hitz & Bieber (2008), have studied the effects of collaborative (team) online knowledge evaluation on the development of learning strategies and contrasted the findings with those of traditional knowledge assessment. They came to the conclusion that group work during online exams increased perceived learning levels and considerably decreased the appearance of surface learning strategies. Paying more attention to oral discussion, Al-Obaydi et al. (2023) studied the effects of online oral reading and other oral activities on EFL college students' cognitive skills such comprehension and class anxiety and they gained very positive results.

Although it is evident that extensive research has been done in the context of a traditional classroom environment on the impact of assessment kinds on the appearance of learning methods (Abduh, 2021; Anderson, 2003; Black & William, 1998; Entwistle, 2000; Ghanbari & Nowroozi, 2021; Rushton, 2005), in the context of online knowledge testing and e-education, the same cannot be said. Only a limited amount of research has been done, leaving room for future studies to confirm or disprove the accuracy of some traditional teaching and assessment-based concepts in the context of the new e-learning environment (Yulianto & Mujtahin, 2021; Zhang et al., 2021; Zlatović et al., 2015). So, this study tries to shed light on essay questions and oral discussion as traditional evaluation techniques to be tested in online education environment.

Method

Design

Generally, there are two main kinds of assessment, each one of them used to measure different level in the process of learning. Formative assessment occurs both before and during the learning process, while summative assessment happens at the conclusion of the learning cycle or at the end of the learning process. Generally, most university teachers, approximately 95%, used MCQ Google form exams and very few of them used essay or other types of exams (5 %). The main reason for that is the ease of use of this type specifically through using of Google Classroom as a main application to teach students during the COVID-19 pandemic. Based on that, this study attempts to test the efficiency of using essay-based exams and oral discussion in monthly formative examinations in comparison with MCQ based tests.

Participants

Two samples were used in the present study. Each consists of thirty EFL college students in The English Department, College of Education for human Sciences at the University of Diyala in Iraq. The participants were third year students and they had half of their subjects online in the academic year 2021-2022. The researchers take every precaution to guarantee that the two groups are more evenly qualified in order to better control the variables that could affect the experiment's outcome. Additionally, the researchers made every effort to limit the impact of unrelated factors that might have an impact on the experiment.

Instrument and data collection

The textbook used for this research study was Larsen Freeman & Anderson (2011) "Techniques and Principals in Language Teaching", which consists of nine chapters. In each chapter, the plan of the experimental group was divided into three lectures per each week and the fourth week is the monthly exam. In each lecture, there is a daily quiz at the end of the lecture to check the students' level of understanding. The quizzes are either just one question or a picture to comment on. At the end of each course, the teacher has at least three-monthly exams and nine daily quizzes. The essential thing is that all these exams are essay questions which require an answer in a short composition, paragraph, or a longer sentence. At the end of the course, before the final examination, both groups were examined orally (in a form of an oral discussion with the tutor of the course) where each student was examined individually by a conducted interview. The final exam of the course is an exam which is a kind of MCQ Google form exam consisting of multiple-choice items or true/false items. The researchers wanted to avoid utilizing Google Form-based examinations for the monthly exams for the experimental group and use it only for the control group in their monthly exams. Figure 1 shows the plan of testing for the two groups and the researchers compared the degrees of the monthly exams of the two groups. The kind of the tests that were used for the experimental group is monthly essay questions in addition to three quizzes for each month. The final mark is out of 20 for each month (2 marks for each quiz, i.e., 6 points altogether, and the monthly essay exam with the maximum number of points 14). The kind of test used for the control group is MCQ Google form exam that consists of

multiple-choice items only without any other testing. The final mark is a number out of 20 for each test, specifically, ten MCQs with two points for each item. Two types of validity were confirmed for the tests used; face validity (by exposing the tests to a jury of experts in the same field) and content validity (by making table of content for the questions and the material used). Both groups received the same teaching as it was conducted by the same tutor and the tutor focused on discussion and questioning trying to aim all language skills during foreign language classes. The identical methodology and topics were used for in the teaching of these two groups and the focus was only on assessing techniques during and after the course.

The chapters of the textbook that were included in this experiment were:

Chapter two: The Grammar Translation Method

Chapter Three: The Direct Method

Chapter Four: The Audio-Lingual Method

Examples of the essay exams are:

Q1. Teaching grammar may be inductive or deductive. What do you prefer? Explain in not more than 100 words.

Q2. Do you believe that a fundamental reason for learning another language is only to be able to read literature written in the target language? Discuss.

Q3. Compare between the goals of teachers in ALM and GTM. What do you prefer?

Q4. What do you prefer to use in teaching English: ALM or Silent way? Explain why.

Q5. Comment on the following statement showing your agreement or disagreement:

Translation is a good teaching technique for beginners.

Examples of MCQs exams are:

1. The feelings of students are not dealt with in.....method.

a-GTM

b-ALM

c-DM

d-ALL of the above

2. Classical method refers to

a- Communicative language teaching

b- Grammar translation method

c- Desuggestopedia

.

d- Silent way

3. The areas of language that are emphasized in GTM were

a-Listening and speaking

b- Reading and writing

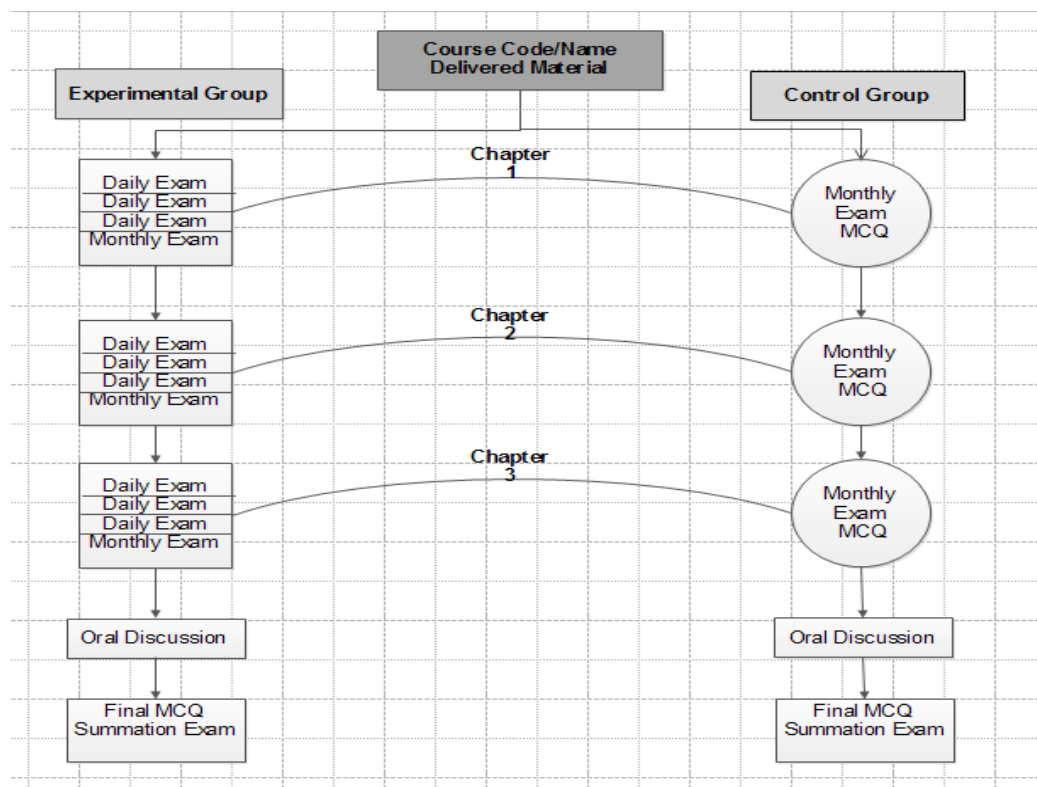
c- Communication and pronunciation

d- Grammar and vocabulary

The final formative exam for the two groups was an oral discussion exam at the end of the course. Each participant was examined individually by the tutor evaluating how they understand the material used during the course. The questioned were much like the essay questions used in the monthly exams but they were used orally and in much depth as the timing of the oral exam were between 20 and 30 minutes for each student. The students were asked to respond orally to the questions by agreeing, disagreeing, explaining, and providing their subjective viewpoints. There were four questions for each student graded by five points for each item. The points rated the students according to their content knowledge, speaking, and pronunciation. All types of questions used in addition to the oral discussion were conducted online in the students' Google class using Google Meet application. All questions were corrected manually by the researchers except for MCQs. Figure 1 explain the marks distribution method within each assessment phase with the two different groups.

Figure 1

the Plan of Testing For The Two Groups



Results

By comparing the marks gained by the two groups in each formative test, the results of each monthly test (particularly, essay exams and MCQs) were calculated to see which of these exams got better results. In each monthly exam, the average of the students in the two tests was calculated, the maximum and the minimum grades of the two groups, and the number of students who passed the two exams.

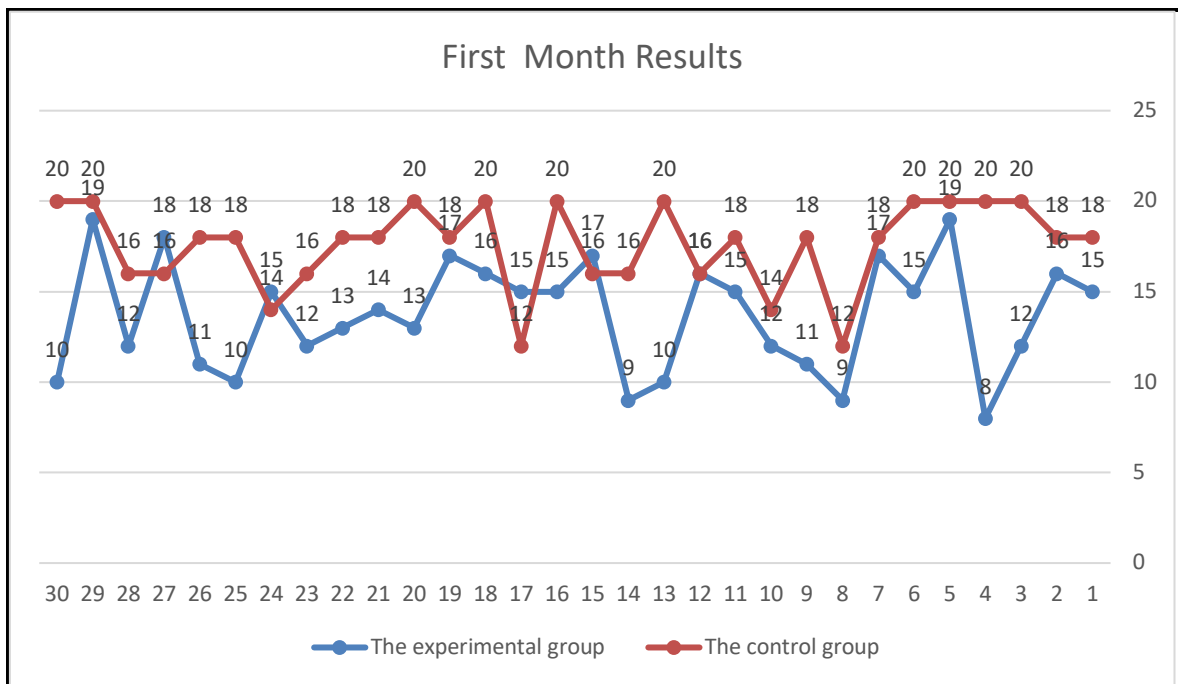
Table 1:

The First Monthly Exam Results

No	The Experimental Group Degree (20%)	The Control Group Degree (20%)
1	15	18
2	16	18
3	12	20
4	8	20
5	19	20
6	15	20
7	17	18
8	9	12
9	11	18
10	12	14
11	15	18
12	16	16
13	10	20
14	9	16
15	17	16
16	15	20
17	15	12
18	16	20
19	17	18
20	13	20
21	14	18
22	13	18
23	12	16
24	15	14
25	10	18
26	11	18
27	18	16
28	12	16
29	19	20
30	10	20
Average	13.7	17.6
Maximum	19	20
Minimum	8	12
No. of Passed	27	30

Figure 2

Line Chart of the First Month Exam Results of Two Groups



According to the results achieved from first month assessment shown in Table 1 and drawing values of Figure 2. It's clear that the control group results are higher than the experimental group. Where the average =17.6 of 30 student who participated. All students passed in the controlled group with using 10 questions with MCQ types of questions. This leads to the rejection of the first hypothesis which states “there is no difference between the experimental group and the control group in the results of the first monthly exam”.

Table 2

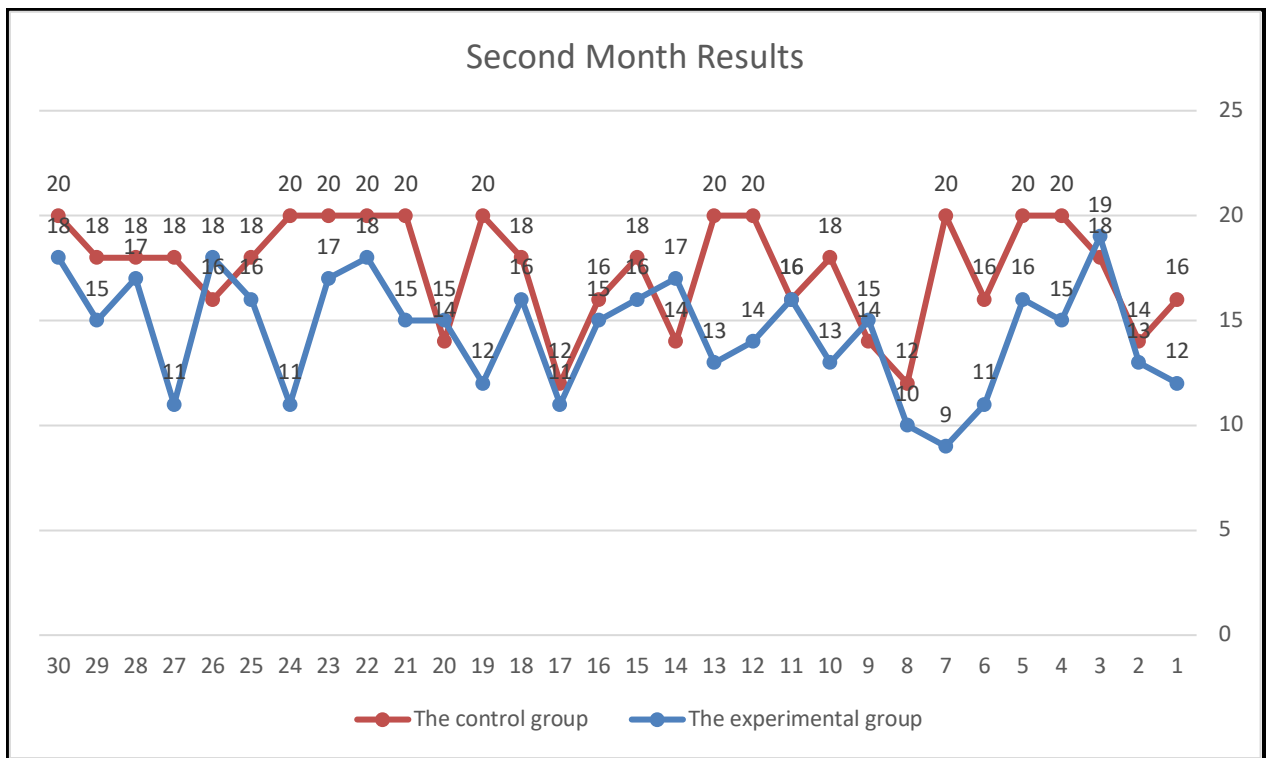
The Second Month Exam Results

No	The Experimental Group Degree (20%)	The Control Group Degree (20%)
1	12	16
2	13	14
3	19	18
4	15	20
5	16	20
6	11	16
7	9	20
8	10	12

9	15	14
10	13	18
11	16	16
12	14	20
13	13	20
14	17	14
15	16	18
16	15	16
17	11	12
18	16	18
19	12	20
20	15	14
21	15	20
22	18	20
23	17	20
24	11	20
25	16	18
26	18	16
27	11	18
28	17	18
29	15	18
30	18	20
Average	14.46667	17.46667
Maximum	19	20
Minimum	9	12
No. of Passed	29	30

Figure 3

Line Chart of the Second Month Exam Results of Two Groups



Further, the results achieved from the second month assessment as they are presented in Table 2 and Figure 3. The findings show that the control group results are significantly higher than the experimental group. Where the average =17.46667 of 30 student who participated. All students passed in the controlled group with using 10 questions with MCQ types of questions. Thus, the second hypothesis is also rejected.

Table 3

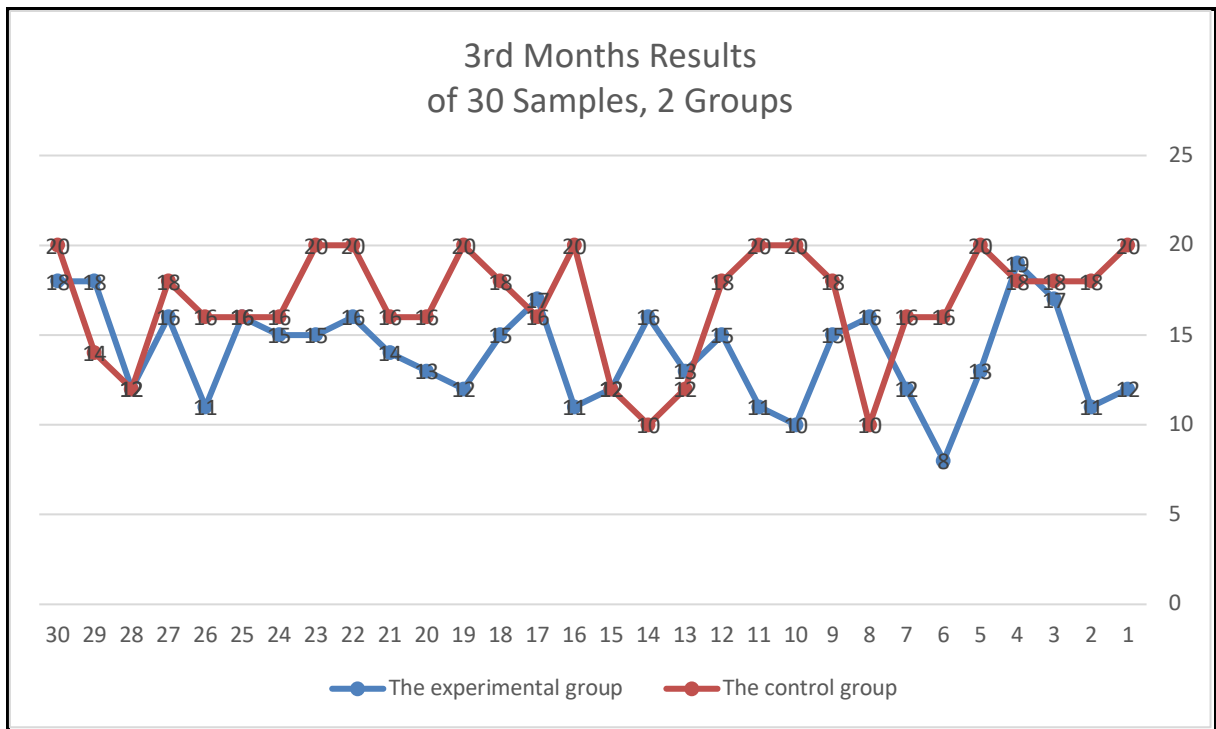
The Third Month Exam Results

No	The Experimental Group Degree (20%)	The Control Group Degree (20%)
1	12	20
2	11	18
3	17	18
4	19	18
5	13	20
6	8	16
7	12	16
8	16	10
9	15	18
10	10	20
11	11	20
12	15	18
13	13	12
14	16	10
15	12	12

16	11	20
17	17	16
18	15	18
19	12	20
20	13	16
21	14	16
22	16	20
23	15	20
24	15	16
25	16	16
26	11	16
27	16	18
28	12	12
29	18	14
30	18	20
Average	13.96667	16.8
Maximum	19	20
Minimum	8	10
No. of Passed	29	30

Figure 4

Line Chart of the Third Month Exam Results of Two Groups



According to the results achieved from the third month assessment as they are presented in Table 3 and Figure 4 the results show that the control group results are continuously

positive and higher than the experimental group. Where the average =17.8 of 30 student who participated. All students passed in the controlled group with using 10 questions with MCQ types of questions. Even in the experimental group the number of the students who passed =29. The third hypothesis is also rejected.

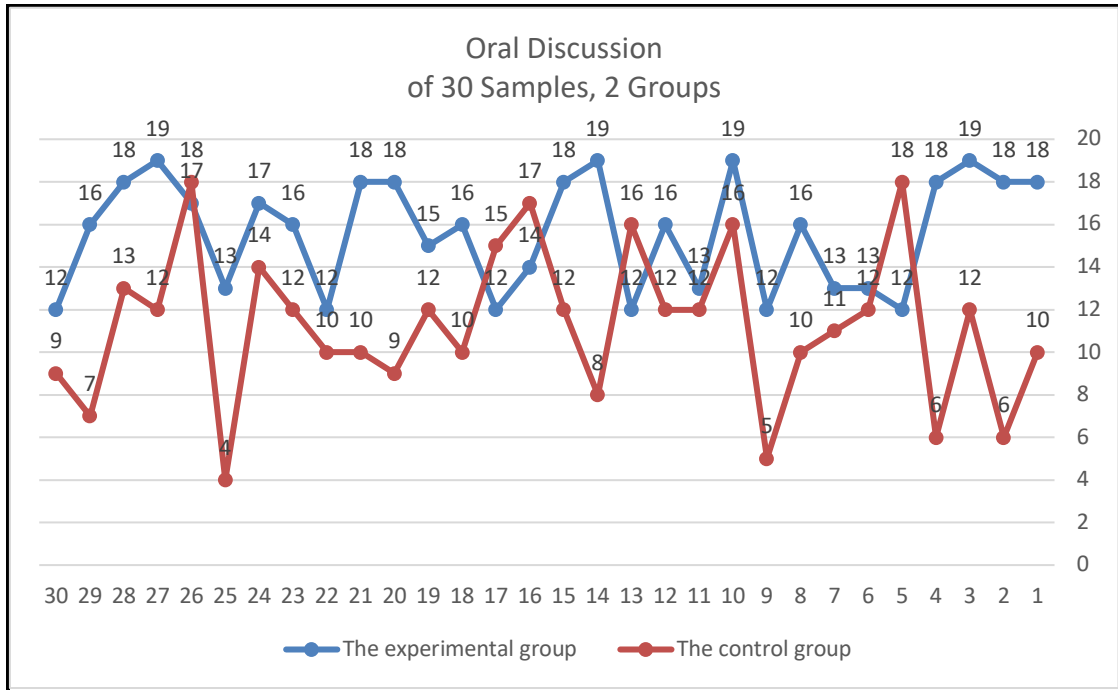
Table 4

The Results of Oral Discussion

No	The Experimental Group Degree (20%)	The Control Group Degree (20%)
1	18	10
2	18	6
3	19	12
4	18	6
5	12	18
6	13	12
7	13	11
8	16	10
9	12	5
10	19	16
11	13	12
12	16	12
13	12	16
14	19	8
15	18	12
16	14	17
17	12	15
18	16	10
19	15	12
20	18	9
21	18	10
22	12	10
23	16	12
24	17	14
25	13	4
26	17	18
27	19	12
28	18	13
29	16	7
30	12	9
Average	15.63333	11.26667
Maximum	19	18
Minimum	12	4
No. of Passed	30	22

Figure 5

Line Chart of The Oral Discussion of Two Groups



The results given from the oral discussion show in Table 4, and Figure 5, explain the overall results that curve up-down and the experimental group achieved the highest results in the average score and in the number of students who passed.

Table 5

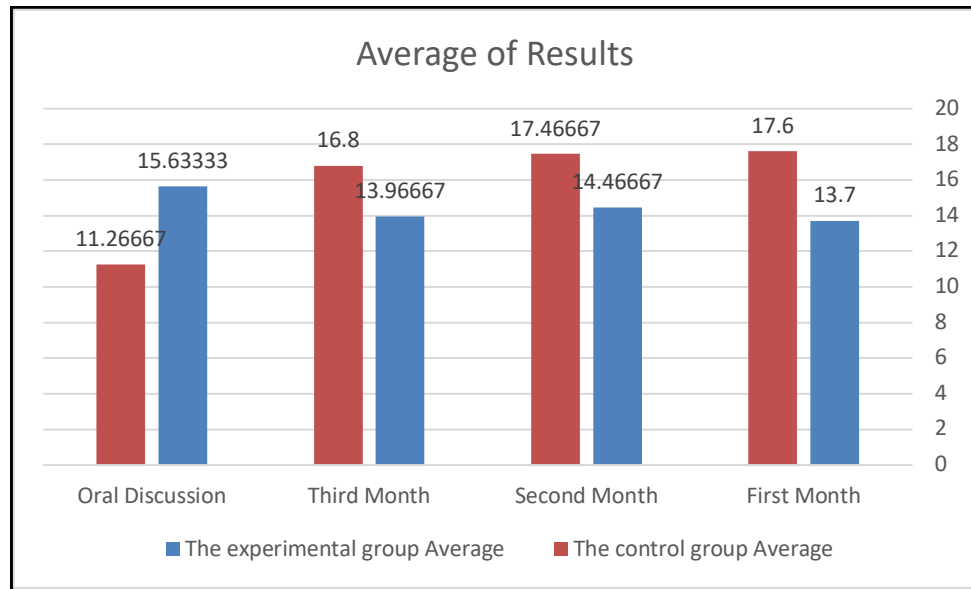
The Average Results of All Course Assessment Types

	The experimental group average	The control group average
First Month	13.7	17.6
Second Month	14.46667	17.46667
Third Month	13.96667	16.8
Oral Discussion	15.63333	11.26667

The average of oral discussion with the experimental group = 15.63333, and the number of students who passed =30. Even though the control group obtained the best results in the previous three assessments but in the oral discussion the students failed in their answers and the minimum assessment marks (4 out of 20). So, the fourth hypothesis is also rejected but in the favor of experimental group.

Figure 6

The Average Assessment Variety of Two Groups

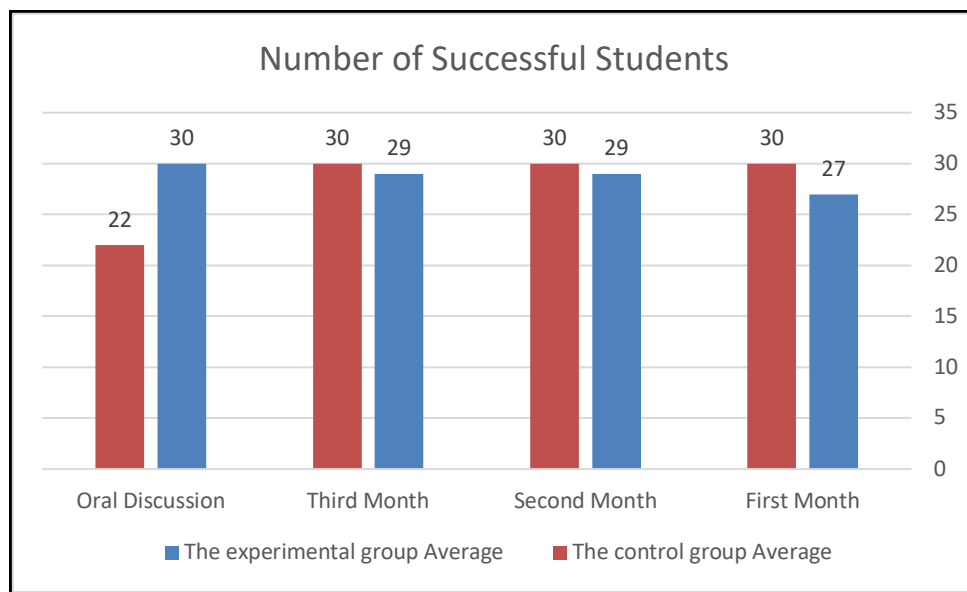
**Table 6**

The Number of Successful Students Results of All Course Assessment Types

	The experimental group average	The control group average
First Month	27	30
Second Month	29	30
Third Month	29	30
Oral Discussion	30	22

Figure 7

Line Chart of the Number of Successful Students Assessment Variety of Two Groups



The results of oral discussion exam (as shown in table 4 and as mentioned by teachers' observation during the exam) proved that the experimental group students, who were tested by using essay exams in their formative examinations and weekly quizzes, gained high levels of cognitive skills and develop deep learning strategies as reflected in their replies in the exams. The questions consist of different techniques such as brainstorming, inference, decision making, and concept acquisition. On the other hand, students in the control group, who were tested only by using MCQ types of questions, failed to cope with the questions asked in the oral discussion exam. Their level of content knowledge was rather low and their overall language skills were much lower than in the experimental group. So, the last hypothesis is also rejected but in favor of the experimental group.

Discussion

As a matter of fact, no one can ignore the difficulties and challenges teachers face in dealing with online assessment and specifically if they want to gain trusted and reliable results that reflect the actual level of students, as supported by Pikhart and Al-Obaydi (2023) and Voinohovska and Doncheva (2022), which is one of the main aims of this study. The verification of the hypothesis of the present study shows that though MCQs can help students, to some extent, to pass the exams, whether formative or summative, they still lack the required content knowledge and the acceptable levels of cognitive skills and learning strategies which are in line with (Marzano & Kendall, 2007; Oliver & Dobele, 2007; Shumway & Harden, 2003; Zlatović et al., 2015). This result based on the fact that the students in the control group showed surface knowledge of the material they supposed to be studied during three months. Unlike the results of control group, the results of the experimental group proves that though the marks of the students are much less than the marks in the control group, but the students levels of understanding, cognitive skills, and learning strategies are much better than the control group as with (Hein & Irvine, 1998; Marzano & Kendall, 2007; Sutadji et al., 2021; Zlatović et al., 2015) who depends only on MCQs in their formative assessment. The main reason for such result is the heavy emphasis on the understanding of the material in the

excremental group trying to satisfy the teacher in the oral discussion test unlike the students of the control group who just pass on the material and depend on guessing in the difficult points.

The specific difficulties college instructor encountered and the conclusions he/she reached after experimenting with various components of the online assessment supported the active involvement of them in managing the educational context in the new online mode by utilizing the resources available which is a fact that no one can ignore as dealt with by Arif (2020). The two types of online assessments that were tested by the present study; MCQ and essay questions were not new and they were used by teachers in traditional teaching environment before the pandemic. These two types are used by teachers online as they are easy to be applied without much training of technological software. It is evident that essay questions put heavy emphasis on teachers specifically online (Abduh, 2021) as they require manual scoring which is time and effort consuming but the result is good and promising. For this reason, most teaches prefer using MCQ for its easiness.

In this vein, the present study's findings indicated that there are still many issues with the Iraqi EFL setting which may be also applied to many other educational contexts in the world, including the need for technological support, instructors' and students' online competency, teacher resources, advanced techniques of online assessment, detecting cheating applications, and learners' technological limitations are still need to be resolved which is in line with Ghanbari and Nowroozi (2021), Hedayati & Marandi (2014), and Tawafak et al. (2023).

The present study provides clear evidence that MCQ types of questions can help students to pass in the exams, and even move to another stage in the college, but without any real understanding to the content knowledge; no development of cognitive skills, and no use of learning strategies (Ghahderijani, et al, 2021; Marzano & Kendall, 2007; Zlatović et al., 2015). This is may be due to many reasons such as guessing, only a little part of the concept is tested, cheating, and the capacity to organize and communicate ideas cannot be measured. The majority of teachers concur that multiple-choice tests are ineffective instruments for evaluating one's capacity to assess and synthesize information or apply knowledge to challenging situations or even figure out problems.

Compared to live (face-to-face) tests and quizzes, students are noticeably more likely to receive answers from others during online exams and quizzes. Keeping the integrity of online exams intact is therefore more difficult (Watson & Sottile, 2010). Detecting and preventing online cheating may be more difficult even though the motivations for cheating in online and offline exams are not noticeably different (Turner & Uludag, 2013). This is due to the fact that, in addition to more conventional cheating techniques, there are also a number of technologies and tools available that can be used to more readily cheat on online tests for instance, using remote desktop and screen sharing, looking up solutions online, and exchanging answers via social media applications. There are many ways to reduce cheating on online exams, including acquiring an offline (face-to-face) proctored exam, creating questions that are difficult for students to

manipulate, lowering the percentage of exam grades that count toward the final course grade, or mixing the MCQs with essay questions.

It is worth mentioning that teaching second or foreign language is not like any other subject. It needs much effort to develop the language and enhancing language skills in addition to the content knowledge. Learning a foreign language online represent a challenge in many pivots to provide ideal leaning context academically and emotionally (Al-Obaydi et al., 2022), so, it is not easy matter to be content with only trivial information about the subject without deep understanding and developing language skills.

The results of the current study added to the body of knowledge regarding the value of employing essay questions and oral discussion to improve content knowledge, cognitive ability, language skills, and learning strategies. The results of this study added a new perspective to the existing empirical literature on the effects of weekly written quizzes and online essay questions specifically with the focus on oral discussion and using different techniques such as brainstorming, inference, decision making, and concept acquisition on the output and performance in second language learning context.

Conclusions

The utilization and use of online tests with multiple-choice questions stimulates a surface learning approach, and a deep learning method is only marginally less aroused. Deep learning strategies and, to a lesser extent, surface learning strategies are encouraged by the introduction and use of online exams in the form of essays. These results can be utilized to develop novel adaptive online assessment techniques that employ different traditional techniques that were previously tested to direct students' learning and builds adaptability system into a series of assessments. So, it is recommended to use a mix of tests types to better evaluate students' performance in learning English language. As recommended by (Fluck, 2019), assessment integrity is crucial for educational institutions since it has an impact on their reputation. To maintain assessment integrity in online exams, it is vital to use standard cheating detection methods in addition to prevention strategies and innovative digital monitoring and validation tools. It also crucial to integrate the aspects of critical pedagogy in EFL contexts (Pikhart et al., 2022) to confirm getting ideal language learning context. Finally, it is recommended that the university administration should continuously train instructors on how to use technological tools in evaluation of the students as well as on the appropriate use of technology in the design and development of the assessment assignments because teachers might encounter problems on a variety of fronts as noted by Ghanbari and Nowroozi ((2021).

Limitations

One of the limitations of the present study is that it was administered in only one university in Iraq, so it is recommended to apply the same idea of the research on more than one university at a time and in different countries. In addition, though using essay questions and oral discussion proved its efficiency in measuring students' content knowledge and cognitive abilities, it still has some demerits such as time consuming

and subjectivity. Thus, it is preferable to work on finding more balanced and practical ways of online assessment to avoid or decrease disadvantages in the two types.

Abbreviations

ELT English Language Teaching

FLL Foreign Language Learning

EFL English as a Foreign Language

MCQ Multiple-choice Questions

Availability of data and material: All data generated by this research are available in the manuscript.

Competing interests: The authors declare no conflict of interests.

Funding: The APC was funded by the University of Hradec Kralove.

Authors' contributions: All authors contributed to the whole manuscript and they all agreed with the final version of the manuscript.

Acknowledgements: The research is a part of the SPEV 2023 project run at the University of Hradec Kralove.

References

- Abduh, M. (2021). Full-time online assessment during COVID-19 lockdown: EFL teachers' perceptions. *Asian EFL Journal*, 28(1), 1–22.
- Al-Obaydi LH, Shakki F, Tawafak RM, Pikhart M and Ugla RL (2023) What I know, what I want to know, what I learned: Activating EFL college students' cognitive, behavioral, and emotional engagement through structured feedback in an online environment. *Front. Psychol.* 13:1083673. <https://doi.org/10.3389/fpsyg.2022.1083673>
- Al-Obaydi, L. H., Pikhart, M., & Derakhshan, A. (2022). A qualitative exploration of emotional intelligence in English as foreign language learning and teaching: evidence from Iraq and the Czech Republic. *Applied Research on English Language*, 11(2), 93-124. <https://doi.org/10.22108/ARE.2022.132551.1850>
- Al-Obaydi, L. H., Rahul, D R & Pikhart, M. (2023). The effect of online oral reading on reading comprehension, reading anxiety, and classroom anxiety among EFL learners. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-023-11950-y>
- Arif, Z. (2020). *Online assessment implemented by English teachers at SMA Al-Islam 1 Surakarta during pandemic COVID-19*. Thesis. English Language Education. Cultures and Languages Faculty. Surakarta, Jawa Tengah: Islamic Institute of Surakarta.

- Aryadoust V, Eckes T and In'nami Y (2021) Editorial: Frontiers in Language Assessment and Testing. *Front. Psychol.* 12:691614. <https://doi.org/10.3389/fpsyg.2021.691614>
- Bridget D. Arend. (2006). *Course Assessment Practice and Student Learning Strategies in Online Courses*. University of Denver.
- Çakmak, F., Namaziandost, E., & Kumar, T. (2021). CALL-enhanced L2 vocabulary learning: using spaced exposure through CALL to enhance L2 vocabulary retention. *Education Research International*, <https://doi.org/10.1155/2021/5848525>
- Fluck, A. E. (2019). An international review of eExam technologies and impact. *Computers & Education*, 132, 1–15. <https://doi.org/10.1016/j.compedu.2018.12.008>
- Forrester, A. (2020). Addressing the challenges of group speaking assessments in the time of the Coronavirus. *International Journal of TESOL Studies*, 2(2), 74–88. <https://doi.org/10.46451/ijts.2020.09.07>.
- Ghahderijani, B. H., Namaziandost, E., Tavakoli, M., Kumar, T., & Magizov, R. (2021). The comparative effect of group dynamic assessment (GDA) and computerized dynamic assessment (C-DA) on Iranian upper-intermediate EFL learners' speaking complexity, accuracy, and fluency (CAF). *Language Testing in Asia*. <https://doi.org/10.1186/s40468-021-00144-3>
- Ghanbari, N., & Nowroozi, S. (2021). The practice of online assessment in an EFL context amidst COVID-19 pandemic: views from teachers. *Language Testing in Asia*. <https://doi.org/10.1186/s40468-021-00143-4>
- Hamadah Alsadoon. (2017). Student's perception of e-assessment at Saudi electronic university. *TOJET: The Turkish Online Journal of Educational Technology*, 16(1), 148.
- Hedayati, H., & Marandi, S. S. (2014). Iranian EFL teachers' perceptions of the difficulties of implementing CALL. *ReCALL*, 26(3), 298–314. <https://doi.org/10.1017/S0958344014000172>
- Hildén, H., & Fröjdendahl, B. (2018). The dawn of assessment literacy – Exploring the conceptions of Finnish student teachers in foreign languages. *Apples – Journal of Applied Language Studies*, 12(1), 1–24. <https://doi.org/10.17011/apples/urn.201802201542>
- Hricko, M., & Howell, S. L. (2006). *Online assessment and measurement*. United States of America: Information Science Publishing.
- Jeanne P. Sewell, et al. (2010). Online assessment strategies: A primer. *MERLOT Journal of Online Learning and Teaching*, 6(1), 297-304. https://jolt.merlot.org/vol6no1/sewell_0310.pdf.
- Khairil and Mokshein. (2018). 21st Century Assessment: Online Assessment. 8(1), 659–672. Retrieved from <https://doi.org/10.6007/IJARBSS/v8-i1/3838>
- Kim, N., Smith, M. J., & Maeng, K. (2008). Assessment in online distance education: a comparison of three online programs at a university. *Online Journal of Distance Learning Administration*, 11(1). [Online] Available: <http://www.westga.edu/~distance/ojdl/spring11/kim111.html>
- Kremmel, B., & Harding, L. (2019). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the

- language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100–120. <https://doi.org/10.1080/15434303.2019.1674855>
- Kremmel, B., Eberharter, K., Holzknicht, F., & Konrad, E. (2018). Fostering language assessment literacy through teacher involvement in high-stakes test development. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (173–194). Cham, Switzerland: Springer.
- Maley, A., & Kiss, T. (2018). *Creativity and English language teaching*. London: Macmillan Publisher.
- Marzano, R. J., & Kendall, J. S. (2007). *The new taxonomy of educational objectives*, 2nd edition. USA: Corwin Press.
- Nowicki, C., & Jones, E. L. (2005). A framework for extending online assessment. *Proceedings of the 6th International Conference on Information Technology Based Higher Education and Training ITHET 2005*, Florida, USA, pp. 12-17.
- Oliver, D., & Dobeles, T. (2007). First year courses in IT: a Bloom rating. *Journal of Information Technology Education*, 6, 347-360. <http://proceedings.informingscience.org/InSITE2007/InSITE07Oliver337.pdf>.
- Pikhart, M., & Al-Obaydi, L. H. (2023). Potential pitfalls of online foreign language teaching from the perspective of the university teachers. *Heliyon*. <https://doi.org/10.1016/j.heliyon.2023.e13732>
- Pikhart, M., Al-Obaydi, L. H., & Tawafak, R. (2022). The implementation of critical pedagogy in digital second language acquisition in higher education: age and gender differences. *Computer-Assisted Language Learning Electronic Journal (CALL-EJ)*, 23(2), 56-75. <http://callej.org/journal/23-2/Pikhart-Al-Obaydi-Tawafak2022.pdf>.
- Robles, M., & Braathen, S. (2002). Online assessment techniques. *Delta Pi Epsilon Journal*, 44(1). https://www.acousticslab.org/dots_sample/module2/RoblesAndBraathen2002.pdf
- Rust, C., O'Donovan, B., & Price, M. (2005). A social constructivist assessment process model: How the research literature shows us this could be best practice. *Assessment and Evaluation in Higher Education*, 30, 231-240. <https://doi.org/10.1080/02602930500063819>.
- Scouler, K. (1998). The influence of assessment method on students' learning approaches: Multiple-choice question examination versus assignment essay. *Higher Education*, 35(4), 453-472. <https://link.springer.com/article/10.1023/A:1003196224280>
- Seifert, T., & Feliks, O. (2018). Assessment & evaluation in higher education online self-assessment and peer-assessment as a tool to enhance student-teachers' assessment skills. *Assessment & Evaluation in Higher Education*, 0(0), 1–17. <https://www.tandfonline.com/doi/abs/10.1080/02602938.2018.1487023>
- Shen, J., Hiltz, S. R., & Bieber, M. (2008). Learning strategies in online collaborative examinations. *IEEE Transactions on Professional Communication*, 51(1), 63-78. DOI: [10.1109/TPC.2007.2000053](https://doi.org/10.1109/TPC.2007.2000053)
- Shumway, J. M., & Harden, R. M. (2003). AMEE guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Medical Teacher*, 25(6), 569-584. DOI: [10.1080/0142159032000151907](https://doi.org/10.1080/0142159032000151907).

- Sutadji, E., Susilo, H., Wibawa, A. P., Jabari, N. A. M., & Rohmad, S. N. (2021). Adaptation strategy of authentic assessment in online learning during the COVID-19 pandemic. *Journal of Physics: Conference Series*, 1810(1), 012059. <https://doi.org/10.1088/1742-6596/1810/1/012059>
- Tawafak RM, Al-Obaydi LH and Pikhart M (2023) Competency categorization and roles of online teachers from the perspective of university students. *Front. Psychol.* 14:1009000. <https://doi.org/10.3389/fpsyg.2023.1009000>
- Turnbull, D., Chugh, R., & Luck, J. (2021). Transitioning to e-Learning during the COVID-19 pandemic: How have higher education institutions responded to the challenge? *Education and Information Technologies*, 1-19. <https://link.springer.com/article/10.1007/s10639-021-10633-w>.
- Turner, S. W., & Uludag, S. (2013). Student perceptions of cheating in online and traditional classes. *Proceedings - Frontiers in Education Conference, FIE*, 1131–1137. <https://doi.org/10.1109/FIE.2013.6685007>
- Voinohovska, V., & Doncheva, J. (2022). Integration of Information and communication technologies in educational theory and practice. In: Proceedings of INTED2022 Conference, 2022, pp. 0452 - 0458, ISBN 978-84-09-37758-9
- Wang, T. H. (2014). Developing an assessment-centered e-Learning system for improving student learning effectiveness. *Computers & Education*, 73, 189-203. <https://www.sciencedirect.com/science/article/abs/pii/S0360131513003266>
- 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). <https://eric.ed.gov/?id=EJ877536>
- Yulianto, D., & Mujtahin, N. M. (2021). Online assessment during COVID-19 pandemic: EFL teachers' perspectives and their practices. *Journal of English Teaching*, 7(2), 229–242. <https://doi.org/10.33541/jet.v7i2.2770>.
- Zhang, C., Yan, X., & Wang, J. (2021). EFL teachers' online assessment practices during the COVID-19 pandemic: Changes and mediating factors. *Asia-Pacific Education Researcher*. <https://doi.org/10.1007/s40299-021-00589-3>.
- Zlatović, M., Balaban, I., & Kermek, D. (2015). Using online assessments to stimulate learning strategies and achievement of learning goals. *Computers & Education*, 91, 32-45. <https://doi.org/10.1016/j.compedu.2015>