The Effectiveness Of Online Assessment Tools Of Al Balqa Applied University Students In Improving Their Language Skills From Their Point Of View

Prof. Mohammad Fawzy Ahmed Bani-Yaseen[1] , Prof. Mohammad Akram Al-Zu'bi[2]

[1]*Basic Science Department, Ajloun University College- Al- Balqa Applied University, Ajloun, Jordan*
mfozy2@gmail.com
[2]*English Language and Literature Department, Ajloun University College, AL-Balqa Applied University, Ajloun, Jordan.*

Abstract—the study aimed at revealing the effectiveness of online assessment tools for Balqa Applied University students in improving their language skills from their perspective; both researchers prepared a questionnaire consisting of (20) items consisting of four areas (listening, reading, writing, and speaking). A triple scale was placed in front of each statement (it applies to me a lot, it sometimes applies to me, it does not apply to me). The study population consisted of students of Balqa University Applied Colleges (Ajlon, Irbid and Al-Huson) in the academic year 2020/2021. The results of the study according to the statistical standard showed that the students' trends were at the intermediate level. It showed a statistical significant difference in favor of Arabic and English compared to non-linguistic disciplines. Also, there was a difference between Arabic and English in favor of English in reading skill. At the skill level, the skill of speaking was ranked first then listening then writing than reading. The results showed that item (6)"The assessment methods measure my ability to understand the content of audible texts" and item (12)" The assessment methods measure my ability to express opinions and the ability to discuss" were ranked first. While item (2)" The assessment methods measure my critical, functional and innovative reading ability" was ranked last.

Index Terms—online assessment tools, students of Balqa Applied University, language skills.

I. Introduction

The methods of distant assessment of students are verified according to their teaching methods. Thus, the traditional teaching that is based on indoctrination requires an assessment method that fits. On other hand, education that goes beyond conservation (moves students from understanding to creativity and moves them from quantity to quality) requires assessment methods that fit in accordance with current and future expectations that relate to the students teaching and assessing them remotely. Distance assessment of students is a structured process of accurate and objective data collection. It requires diversity in methods according to specific targets to reach estimates used in decision-making.

The teacher's decision to assess students remotely is cumulative. There is a prior assessment to determine the level of students. Also, there is a formative assessment that seeks improvement and development. Furthermore, there is a diagnostic assessment that aims to identify strengths and weaknesses. Finally, there is a final assessment aimed at determining the extent of achieving the educational goals for expected inputs and outputs. To achieve that, teachers are allowed to use easy and objective electronic tools in the distant assessment that provide graphs and statistical data. There are some warns for the remote assessment of students like the truth of data e.g. cheating. The findings of Almenie (2018) study showed a high level of cheating in the distance education.

There are practices and methods for assessing students that can be employed and developed in proportion to distance learning, for example, open book examinations (Basyouni, 2016). Open book

examinations can assess the advanced level of skills and thinking. But one of its caveats is students' belief that there is no need for prior study. One of the caveats is the belief that students do not need to study earlier. To treat this problem, examples of this type of question can be provided and model answers provided before students arrive at the actual evaluation phase. The Turnitin tool can be employed to be the starting and ending points by all students like using E-mail.

Students can be assessed with essay questions. They are comprehensive, and they can cover the purposeful content (Basyouni, 2016). One of its caveats in the assessment is to draw attention to the effort in its preparations as well as poor printing speed skills for students. Students can write the answers by using their hands and download them after training them. Moreover, the students can be assessed by short examinations. They are classified into different types: short answers, true and false, and multiple choice. They are difficult to prepare. A bank of questions can be created and distributed randomly to address the problem of communication between students during the exam. Besides, it's not possible to go back to the question after answering and selecting with the appropriate time for each accurately.

Fragmented tests are the most popular in distant student assessment. They are rested two or three tests: the first exam shall be allocated to (25) degrees, the second exam shall be allocated to (25) degrees, and 50 degrees shall be allocated to the final test. The students can interact with their teacher by using the PowerPoint presentation with sound and image and the dialogue is direct (Almenie, 2018). Thus, the instruction is based on learner-centered approach and the assessment will be more accurate. However, the teacher must be aware of the different available electronic possibilities to students. Students can be asked to achieve a project. Also, they can prepare a file but the teachers should pay attention on the possibility that students will be dependent on others in preparing these projects.

The assessor can employ some tools that measure students' rate of quotation, for example, "Safe Assign". It is an effective assessment tool for students' researches and projects (Alothman, 2020). There are also e-evaluation applications called "One Note" to set up the achievement file and cloud storage towards "Google Drive-One Drive-Dropboxis" to transfer information among students easily. Besides, there are Assessment methods provide an opportunity to provide feedback electronically like using "Google Forms-Class Maker-Edmodo". It is required from the teacher who assesses his students an appropriate planning and rigorous implementation. Besides, there are requirements include teacher, students and society.

To talk about those requirements that respond to the implications of the Cognitive and technological revolution in general and the Corona pandemic in particular, studies have been conducted, including:

Megdadi (2020) conducted a study that aimed at revealing perceptions of high school students in government schools in Jordan to use distance education in light of the Corona crisis and its developments. After selecting a sample of (167) Male and female students from government schools in the Kasbah District of Irbid and applying the descriptive survey Method, the study revealed that there is a positive impact of the use of distance education in light of the new crisis in Kurna. The researcher encouraged other researchers to carry out more studies on different samples and stages and provide adequate support to the Ministry of Education to manage distance Education

Almeqren (2019) investigated a study aimed at identifying the effect of e-learning by using "Edmodo" on the achievement of the computer course at cognitive levels (remembering, understanding) at the female students of second secondary grade in Riyadh, and the trends towards it. The study sample consisted of 62 female students at the first secondary grade in Riyadh, in the second semester 1437/ 1438 AH. They were divided into experimental and control groups. After implementation of a set of tools and materials according to a website on Edmodo network that includes an e-lesson, an assignment, a test, and activities that have been designed by the researcher, the obtained Data was analyzed. The findings of the study showed statistically significant differences between the average of the experimental and control groups in favor of the experimental group.

Bani-Yasain (2018) conducted a study that aimed at uncovering the attitudes of the students at Balqa Applied University towards their evaluation methods. To achieve the aim of the study, the researcher designed a measurement of (23) items, distributed on 150 students from Ajloun University

college and 10 teachers in the academic year 2016 /2017. The most important result showed that all the fields achieved the intermediate level. The field of tests ranked first, the field of assessment methods generally ranked second, and the field of quarterly work and class discussions ranked third.

Bani-Yaseen (2016) conducted a study aimed at analyzing the General Secondary examination (GSE) questions of Arabic language subject in Jordan according to Bloom's cognitive level. The study concluded that, all questions in summer and winter session focus respectively, on understanding, application, remembering, application, and analysis questions. Furthermore, the assessment questions have been neglected with (0%).

Bani-Yaseen et al. (2013) carried out a study aimed at investigating the advantages of an evaluation style in improving the lingual knowledge of (45) females' students at Ajloun College. The researchers distributed the sample of the study into three experimental groups according to the ways of evaluation randomly. Fifteen students were evaluated by the teacher without discussing the test after evaluation, fifteen students were evaluated by classmates; and fifteen students was assessed by themselves. The finding of the study showed that there were significant differences among the performance means due to the group that was assessed by classmates and themselves.

Alshaikhi, (2015) conducted a study aimed at revealing the role of the university professor in improving the quality of the methods of evaluation. This study contributed in improving the learning and teaching process in universities. The study sample consisted of 60 instructors in Umm-Alqura and King Faisal Universities. The study concluded that teaching staff could practise appropriate evaluation methods that would contribute to good quality.

Suleiman (2010) conducted a study aimed at evaluating the practices of the staff at a University in assessing their students' achievement according to some variables. The results of the study showed that there were many advantages in such practices, including: a variety of questions, and providing model answers after students hand over the results of their tests. For the negatives, they were mentioned by the students as follows: Lack of accuracy and objectivity in monitoring research or report scores and failure to follow students during preparing research.

Abu Nimreh, M.K. & Al Hadidi (2009) conducted a study aimed at revealing the training needs of members of the staff in the educational science faculties of Jordanian private universities from their point of view. The results of the study showed that teachers need to be trained to prepare tests and analyze results was moderate.

Aljabouri (2008) conducted a study aimed at evaluating the performance of the university professor in the light of professional and personal qualities from the perspective of graduate students at the University of Babylon. To achieve that, the researcher distributed a questionnaire to (100) Male and female students. The questionnaire consisted of two scopes: the professional qualities scope consisting of (30) items and the personal qualities consisting of (12) items. The results of the study revealed that the scope of evaluation obtained a low level.

Burqa'an A. & Alrabea' (2003) conducted a study aimed at evaluating the teaching practices of educational concepts from the students' perspective. The researcher distributed a questionnaire to (350) Male and female students. The study showed that the methods of evaluation used do not cover the objectives of the course. Also, the questions of the examinations were unclear.

Aljado' (1999) conducted a study aimed at detecting teaching and evaluation strategies for development. The study showed that there are several problems facing female students and teachers. Those problems are: the reliance of female students on literal preservation, loss of confidence in teacher assessment, the need to adopt quality testing criteria, and the diversification of evaluation formats.0

This present study is consistent with the previous studies that focus on assessment. The Megdadi study (2020) revealed a significant positive impact of distance learning from the perspective of students. The results of the study conducted by (Almekren, 2019) also showed a positive impact on the achievement of the distance learning group, as well as the results of the Suleiman and Khalid study (2010), showed that there are many positives in faculty staff practices while assessing their students.

The Alshaikhi, (2015) study (2015) revealed that teaching staff can exercise appropriate evaluation methods that contribute to improving the output of education. The results of Abu Namra and Al-Hadidi study (2009) showed the need for teachers to be

trained in the preparation of tests. Studies such as the Al-Jubouri study (2008), which aimed to evaluate the performance of the university professor, showed that the field of the evaluation was at a low level. The results of the Burqa'an A. & Alrabea' (2003) study also showed that the evaluation methods used did not cover the course's objectives, and that the examination questions were unclear. The results of Al-Jadou' and Hasan Study (2001) showed a lack of confidence in teacher assessment and the need to adopt quality testing criteria, and the diversification of assessment formats.

The results of the Bani-Yaseen study (2018) showed that the field of testing was ranked first, the field of evaluation methods generally ranked second, and the area of semester work and class discussions ranked third. The study of Bani Yassin (2016), which aimed to analyze the questions of the General Secondary School Examination of the Arabic Language Course in Jordan according to Bloom's cognitive levels, showed that the level of recollection was highest, while the level of the evaluation was the lowest.

The current study was distinguished in its area of research. It tries to reveal the effectiveness of the students' assessment methods at Al-Balqa Applied University in developing their lingual skills from their point of view. The study has taken a preliminary idea from some of the previous studies' tools. It works to develop them for the purposes of the current study and works to formulate the problem of the study.

A. The problem of the study

The assessment of students is one of the core functions of a member of the university faculty. It is expected to verify the assessment methods and choose the most appropriate ones. In the Korna pandemic, the researcher observed that educational institutions, primarily universities, were more concerned with distance education than with the methods of assessment in general. As well as researchers fail in the field of distance evaluation methods to develop their language skills, namely, reading, writing, listening, and reading skills. The language center in Jordan University has claimed that the sudden transfer in teaching language caused a shock between the students and their teachers. Besides, there are challenges for students, for teachers and for the nature of language teaching and learning.

So, recognizing the reality of distance assessment clearly enables the ability to diagnose, and develop assessment-related plans to achieve development. Besides, the effectiveness of student assessment methods in linguistic development is a fundamental foundation of the educational process so the current study tries to answer the following question:

Are there statistically significant differences at (α = 0.05) in the effectiveness of the distance assessment methods of the students at Al-Balqa Applied University in developing their language skills from their point of view due to the variables of specialization (Arabic, English, etc.)?

B. Objectives of the study

The aim of this study is to achieve multiple objectives, including:
- Detecting the effectiveness of distance assessment methods for students in developing their language skills.
- Revealing the importance of assessment methods in general and the relationship of assessment to education in particular.
- Making recommendations to educational institutions.

C. The importance of the study

- Theoretical significance: The importance of this study in its research field is exceptional. Most teachers relied on the assessment methods they had with their students prior to the Corona pandemic. So, the educational institutions lack of interest of in conducting studies that revealed the effectiveness of modern evaluation methods concurrent with distance learning.
- Practical importance: The current study seeks results that reveal the effectiveness of distance assessment methods for students in developing their language skills.

D. The limitations of the study

- Time limitation: Only summer semester 2020-2021.
- Spatial limitation: Only northern colleges (Irbid College, Huson College, and Ajlon College).
- The human limitation: Limited to bachelor students.
- Substantive limitation; Limited to assessment methods according to the study tool.

E. The operational definitions

- Assessment methods: The procedures practiced by the professor for determining students' grades.

-Language skills: Major language skills (reading, writing, listening and conversation) and their minor skills, represented the behavioral indicators.

- Distance assessment: Students' performance on the test provided to them.

(Student residence)

- Al-Balqa Applied University Students: Bachelor's students who study at Northern Colleges (Irbid College, Huson College, and Ajlon College).

II. METHODS AND PROCEDURES

The researchers surveyed the theoretical literature and the previous studies relating to the methods of remote assessment from their point of view. They identified the methodology of the study, the tool of the study (reliability and validity), the population of the study, the sample of the study, as well as the variables of the study and the appropriate statistical treatments.

A. Methodology of the study

The current study adopted the quantitative approach. The analytical descriptive approach aims to describe the phenomenon and to determine its reality to reach the scientific conclusions that can be discussed. To achieve the purpose of the study, the researcher built the study instrument which is a questionnaire that identified language skills (reading, writing, listening, and speaking). It consisted of 20 items (five statements or each skill).

B. Study variables

The study examined the following variables: First: the independent variable which is "the effectiveness of online assessment tools of Al Balqa Applied University students".

Second: a dependent variable which is" improving their language skills from their point of view".

C. Statistical standard
The Likert triple scale was adopted to correct the study instruments. It gave each item one grade (large, medium, low) which was represented digitally (3, 2, and 1), respectively. The following

scale was adopted for the purpose of analyzing the results:

1.00-1.66 Low
1.67-2.33 Medium
2.34- 3.00 Large

The scale has been calculated using the following equation:

Maximum Scale (3) - Minimum Scale (1)

_____

Number of required categories (3)

.66=   3-1
_____

3

Then adding the answer (0.66) to the end of each category.

D. The validity of the instrument

The researchers relied on the content validity for the study instrument which consisted of a list of major language skills (reading, writing, listening, and speaking), and the minor sub-skills to identify the behavioral indicators for each sub-skill. The instrument was presented to a number of judgers specialized in language curricula, teaching methods, measurement, and evaluation. They were asked to express their views on the items of the instrument in terms of belonging to the subset skills of the target skill, their clarity, their language integrity, as well as their suitability for the study sample, and any other observations they deem appropriate. The judgers' observations, which included an adjustment to some of the sub-skills were followed until the final version represented in (schedule 10).

E. The validity of the scale building

In order to extract the indicators of the scale validity, the correlation coefficients of each item and the total degree were extracted. Also, the correlation coefficient between each item and its relation to the domain belonging to, and between the domains and the total degree were extracted in a survey sample from outside the sample consisting of 30. The correlation coefficients of the items with the instrument as a whole ranged from (0.85 to 0.89), and to the field (from 0.68 to 0.92) (see table 1).

TABLE 1 CORRELATION COEFFICIENTS BETWEEN THE STATEMENT AND THE TOTAL DEGREE AND THE DOMAIN TO WHICH IT BELONGS

Item No.	Correlation coefficients with the domain	Correlation coefficients with the instrument	Item No.	Correlation coefficients with the domain	Correlation coefficients with the instrument	Item No.	Correlation coefficients with the domain	Correlation coefficients with the instrument
1	.89**	.85**	8	.87**	.78**	15	.86**	.85**
2	.68**	.58**	9	.88**	.79**	16	.84**	.73**
3	.83**	.68**	10	.92**	.81**	17	.84**	.77**
4	.79**	.67**	11	.83**	.68**	18	.86**	.77**
5	.80**	.79**	12	.77**	.72**	19	.91**	.88**
6	.82**	.83**	13	.92**	.87**	20	.90**	.89**
7	.92**	.84**	14	.85**	.86**			

* Statistically significant at (0.05).

** Statistically significant at (0.01).
It should be noted that all correlation factors were of acceptable grade and statistically significant therefore none of these items has been deleted. The correlation coefficient of the domain with the total degree has been extracted, and the correlation coefficients between the domains (see table 2).

TABLE 2 THE CORRELATION COEFFICIENTS BETWEEN THE DOMAINS WITH THE TOTAL DEGREE

	Reading	Listening	Speaking	writing	The total degree
Reading	1				
Listening	.756**	1			
Speaking	.768**	.854**	1		
Writing	.806**	.766**	.831**	1	
The total degree	.901**	.918**	.940**	.924**	1

* Statistically significant at (0.05).

** Statistically significant at (0.01).

Table (2) shows that all correlation coefficients were of acceptable grade and statistically significant which indicates an appropriate degree of construction validity.

F. reliability of the instrument

In order to ascertain the reliability of the study instrument, the (test-retest) was verified by applying the scale. Then, reapplying the scale on out-of-sample group of 30 two weeks later. After that, Pearson correlation factor between participants' estimates in both times was computed. The constancy factor has also been calculated in the way of internal consistency by the Cronbach-alpha equation. Table 3 shows the internal consistency factor according to the Cronbach-alpha equation, the test-retest reliability of the domains and the total degree. These values were considered appropriate for the purposes of this study.

TABLE 3 INTERNAL CONSISTENCY COEFFICIENT (CRONBACH ALPHA) THE TEST-RETEST RELIABILITY OF THE DOMAINS AND THE TOTAL DEGREE

Domain	test-retest reliability	Internal consistency coefficient
Reading	0.92	0.85
Listening	0.89	0.82

Speaking	0.90	0.89
Writing	0.91	0.90
The total degree	0.92	0.91

G. The study population and its sample

The study population was made up of students from Al- Balqa Applied University who hold a bachelor's degree. They were selected from three colleges (Ajloun University College, Al-Huson College and Irbid College). The questionnaire has been distributed to students in these colleges. The researchers selected (130) responses for Arabic students, (130) for English students and (130) for those who are not specialized in Arabic and English students according to the Table (4).

TABLE 4 FREQUENCIES AND PERCENTAGES ACCORDING TO THE STUDY VARIABLES

	Categories	Frequencies	Percentage
Specialization	English	130	33.3
	Arabic	130	33.3
	Other	130	33.3
	Total	390	100.0

III. RESULTS AND DISCUSSION OF THE STUDY

The researchers examined the results of the study in a way that highlighted the most important findings with explanations. They relied on the theoretical literature and the related studies. They showed the main areas of agreement and disagreement with those studies, and suggested recommendations according to the following study question:

Are there statistically significant differences at (α = 0.05) in the effectiveness of the distance assessment methods of the students at Al-Balqa Applied University in developing their language skills from their point of view due to the specialization (Arabic, English, etc.)?

To answer this question, the mean scores and standard deviations of the effectiveness of the distance assessment methods of the students at Al-Balqa Applied University in developing their language skills have been extracted from their point of view, as shown in the table (5).

TABLE 5 THE MEAN SCORES AND STANDARD DEVIATIONS OF THE EFFECTIVENESS OF THE DISTANCE ASSESSMENT METHODS OF THE STUDENTS AT AL-BALQA APPLIED UNIVERSITY IN DEVELOPING THEIR LANGUAGE SKILLS HAVE BEEN EXTRACTED FROM THEIR POINT OF VIEW IN DESCENDING ORDER BY MEAN SCORES

Rank	No.	Domain	Mean-scores	standard deviations	Level
1	3	Speaking	2.28	0.85	Average
2	2	Listening	2.26	0.82	Average
3	4	Writing	2.20	0.89	Average
4	1	Reading	2.19	0.90	Average
		The total degree	2.23	0.91	Average

Table 5 shows that the mean scores ranged from (2.19 to 2.28). The first skill is speaking then listening, then writing, then reading. The mean score of the effectiveness of the distance assessment methods of the students at Al-Balqa Applied University in developing their language skills as a whole was an average (2.23).

There has been a difference between the mean scores of speaking and listening skills on the one hand and the skills of reading and writing on the other hand. This difference is possibly due to the sudden shift from face education to distance education. The students found oral communication more easily than written communication, particularly indirect and immediate feedback in its individual and collective types. For written feedback, it may be delayed because of the teacher's interaction in oral communication being more than written communication, especially for a teacher whose level

of interaction in education has not been higher than the standards of quality.

An increase in the mean score of speaking skill is observed above listening. It is possibly due to students' care to obtain a degree of participation. Part of the students' assessment was devoted to their interventions, interaction and oral presentations. The progress of the mean scores in writing skill over reading skill despite of its difficulty may be attributed to the fact the item (20) which states that "The methods of assessment measure my ability to take into account the written task which is assigned". So, it was ranked second as a general written skill that language teachers and others focused on. The clarity of the test items is a priority whether they are constructive or objective.

The mean scores and standard deviations of the study sample members' estimates were calculated on a skill-by-skill basis, as follows:

1. Reading skill

TABLE 6 THE MEAN SCORES AND STANDARD DEVIATIONS FOR READING SKILL RANKED DOWNWARD ACCORDING TO THE MEAN SCORES

Rank	No.	Items	Mean-scores	standard deviations	Level
1	3	The assessment methods measure my ability to induce and summarize ideas.	2.24	.764	Average
2	5	The assessment methods measure my ability to detect and emulate the beauty aspects of reading text	2.20	.766	Average
3	1	The assessment methods measure my ability to read various poetry and anthology texts, both Arabic and translated.	2.19	.794	Average
4	4	The assessment methods measure my ability to To identify ideas and add to them, and to address its problems.	2.19	.747	Average
5	2	The assessment methods measure my critical, functional and innovative reading ability	2.10	.755	Average

Rank	No.	Items	Mean-scores	standard deviations	Level
		Reading skill	2.19	.676	Average

Table 6 shows that the mean scores ranged from (2.10 to 2.24). Item (3), which states that "The assessment methods measure my ability to induce and summarize ideas." ranked first and with mean score of (2.24). That is possibly because the induction and summary are general skills that all teachers seek to achieve. This includes both linguistic and non-linguistic disciplines.

Item (2), which reads "The assessment methods measure my critical, functional and innovative reading ability" ranked last with a mean score of (2.10). That is possibly because of the nature of the questions. Most of the questions are objective and do not measure the level of critical, functional and innovative reading, as well as the lack of attention of non-linguistic teachers to measure the advanced levels of reading skill as an accurate language skill. The language teacher considers reading as a purpose while it is a means for non-linguists. The mean score of reading skill as a whole was (2.19).

2. Listening skill

TABLE 7 THE MEAN SCORES AND STANDARD DEVIATIONS FOR LISTENING SKILL RANKED DOWNWARD ACCORDING TO THE MEAN SCORES

Rank	No.	Items	Mean-scores	standard deviations	Level
1	6	The assessment methods measure my ability to understand the content of audible texts.	2.33	.762	Average
2	8	The assessment methods measure my ability to taste audible texts, and to show the beauty elements in.	2.28	.764	Average
3	7	The assessment methods measure my ability to distinguish the compound relationships in audible texts.	2.25	.741	Average
4	9	The assessment methods measure my ability to apply audible texts in my life, analyze them and draw insight into them.	2.25	.751	Average
5	10	The assessment methods measure my ability to apply critical criteria in criticizing audible texts with reasoning, argument and proof.	2.18	.775	Average
		Listening skill	2.26	.672	Average

Table 7 shows that the mean scores ranged from (2.33to 2.18). Item (6), which states that "The assessment methods measure my ability to understand the content of audible texts." ranked first with a mean score of (2.33). The reason may be that distance learning has increased the students' focus on relying on the listening skill to understand the content of lectures, as well as the accessibility and listening to recorded lectures available to students at any time.

Item (10), which reads "The assessment methods measure my ability to apply critical criteria in criticizing audible texts with reasoning, argument, and proof." ranked last with mean score of (2.18). That is possible because of the nature of the questions. Most of the questions are objective of short answers. These questions do not give students the opportunity to apply the criticism criteria for what they hear. The mean score of listening skill as a whole was (2.26).

3. Speaking skill

TABLE 8 THE MEAN SCORES AND STANDARD DEVIATIONS FOR SPEAKING SKILL RANKED DOWNWARD ACCORDING TO THE MEAN SCORES

Rank	No.	Items	Mean-scores	standard deviations	Level
1	12	The assessment methods measure my ability to express opinions and the ability to discuss.	2.33	.753	Average
2	11	The assessment methods measure my ability to speak boldly and fluently in the sound language.	2.29	.767	Average
3	13	The assessment methods measure my ability to present ideas in sequence.	2.29	.745	Average
4	15	The assessment methods measure my ability to propose solutions to a variety of problems.	2.28	.772	Average
5	14	The assessment methods measure my ability to argue with mental proof.	2.19	.781	Average
		Speaking skill	2.28	.690	Average

Table (8) shows that the mean scores ranged from (2.19 to 2.33). Item (12), which states that "The assessment methods measure my ability to express opinions and the ability to discuss." ranked first and with a mean score of (2.33). This may be due to the interest of teachers during assessing the class's work (30 marks) for the students to stimulate them to

discuss subjects. Also, the language teachers focus on oral presentations according to linguistic criteria.

Item (14), which reads "The assessment methods measure my ability to argue with mental proof." ranked last with a mean score of (2.19). To some extent, this meets with the item (10) in the skill of listening. The nature of the objective and essay questions with short answers may not give students

the opportunity to apply critical criteria so they could not present the transcript or mental evidence while speaking. Besides, student numbers and technical problems may prevent teachers from addressing conversation skills at advanced levels. The mean score for speaking skill as a whole was (2.28).

4.   Writing skill

TABLE 9 THE MEAN SCORES AND STANDARD DEVIATIONS FOR WRITING SKILL RANKED DOWNWARD ACCORDING TO THE MEAN SCORES

Rank	No.	Items	Mean-scores	standard deviations	Level
1	20	The assessment methods measure my ability to take into account the assigned written task.	2.31	.759	Average
2	19	The assessment methods measure my ability to analyze, discuss and link ideas and textual relationships	2.21	.807	Average
3	18	The assessment methods measure my ability to employ a system of morphological and syntactical structure and grammar.	2.18	.806	Average
4	17	The assessment methods measure my ability to write functional writing.	2.16	.782	Average
5	16	The assessment methods measure my ability to write critical and creative writing.	2.13	.803	Average
		Writing skill	2.20	.717	Average

Table (9) shows that the mean scores ranged from (2.13to 2.31). Item (20), which states that "The assessment methods measure my ability to take into account the assigned written task." ranked first and with a mean score of (2.31). This may be due to the interest of teachers to make writing task a high degree of clarity.

Item (16), which reads "The assessment methods measure my ability to write critical and creative writing." ranked last with a mean score of (2.13). This is probably because the tests probably do not include this level of questions. The mean score for speaking skill as a whole was (2.20).

TABLE 10 THE MEAN SCORES AND STANDARD DEVIATIONS OF THE EFFECTIVENESS OF THE DISTANCE ASSESSMENT METHODS OF THE STUDENTS AT AL-BALQA APPLIED UNIVERSITY IN DEVELOPING THEIR LANGUAGE SKILLS HAVE BEEN EXTRACTED FROM THEIR POINT OF VIEW IN DESCENDING ORDER BY MEAN SCORES

No.	language skill	Mean-scores	standard deviations	Rank

9	The assessment methods measure my ability to read various poetry and anthology texts, both Arabic and translated.	2.19	.794	9
13	The assessment methods measure my critical, functional and innovative reading ability	2.10	.755	13
6	The assessment methods measure my ability to induce and summarize ideas.	2.24	.764	6
9	The assessment methods measure my ability to To identify ideas and add to them, and to address its problems.	2.19	.747	9
8	The assessment methods measure my ability to detect and emulate the beauty aspects of reading text	2.20	.766	8
1	The assessment methods measure my ability to understand the content of audible texts.	2.33	.762	1
5	The assessment methods measure my ability to distinguish the compound relationships in audible texts.	2.25	.741	5
4	The assessment methods measure my ability to taste audible texts, and to show the beauty elements in.	2.28	.764	4
5	The assessment methods measure my ability to apply audible texts in my life, analyze them and draw insight into them.	2.25	.751	5
10	The assessment methods measure my ability to apply critical criteria in criticizing audible texts with reasoning, argument and proof.	2.18	.775	10
3	The assessment methods measure my ability to speak boldly and fluently in sound language.	2.29	.767	3
1	The assessment methods measure my ability to express opinions and the ability to discuss.	2.33	.753	1
3	The assessment methods measure my ability to present ideas in sequence.	2.29	.745	3
9	The assessment methods measure my ability to argue with mental proof.	2.19	.781	9
4	The assessment methods measure my ability to propose solutions to a variety of problems.	2.28	.772	4
12	The assessment methods measure my ability to write critical and creative writing.	2.13	.803	12
11	The assessment methods measure my ability to write functional writing.	2.16	.782	11

10	The assessment methods measure my ability to employ a system of morphological and syntactical structure and grammar.	2.18	.806	10
7	The assessment methods measure my ability to analyze, discuss and link ideas and textual relationships	2.21	.807	7
2	The assessment methods measure my ability to take into account the assigned written task.	2.31	.759	2

Table 10 shows that Items (6) and (12), which state respectively that " The assessment methods measure my ability to understand the content of audible texts and the assessment methods measure my ability to express opinions and the ability to discuss." The first item is considered as listening skill and the second one considered a speaking skill. That is possible because the teacher and the student prefer the oral communication represented by listening and speaking to communicate easily rather than editorial communication represented by reading and writing. Item (20), which states "The assessment methods measure my ability to take into account the assigned written task." ranked second with a mean score of (2.31). This may be due to the concern of all teachers to make writing task too clear because the teacher is not with the students during the test. The third higher items were (11 and 13). They state respectively that "The assessment methods measure my ability to speak boldly and fluently in sound language" and "the assessment methods measure my ability to present ideas in sequence". That is possible because the effectiveness of oral communication distance learning comes before reading and writing.

Item (16) "The assessment methods measure my ability to write critical and creative writing" has the penultimate ranking with an average of (2.13). Item (2) which stated "The assessment methods measure my critical, functional and innovative reading ability" ranked the last one. This may be due to the fact that the tests are likely to be free of this level of questions and the correction of the tests was electronic. As well as those two skills are considered editorial and advanced skills.

The results of the study agree with Bani-Yasain Study (2016), which showed the highest level of remembrance, while the evaluation level was the last. Burqa'an A. & Alrabea' study (2003) revealed that the evaluation methods used do not cover the course goals. This is consistent with the results of the Bani-Yasain study (2018). The results were moderate with the studies of (Suleiman (2010), 2010; Alshaikhi, 2015; Almeqren, 2019). The current study disagrees with the Megdadi (2020) which showed a significant positive impact of distance education. The results of Aljado' study (2001) showed a loss of confidence in teachers' evaluation, the need to adopt quality testing standards, and the variation of evaluation forms. While the results of the Abu Nimreh & Al Hadidi, (2009) study showed the need for teachers to be trained on the preparation of tests and Aljabouri (2008) showed a low level of evaluation.

To show the effectiveness of the distance assessment methods of the students at Al-Balqa Applied University in developing their language skills from their point of view according to their specialties, mean scores and standard deviations have been derived as shown in table (11).

TABLE 11 THE MEAN SCORES AND STANDARD DEVIATIONS OF THE EFFECTIVENESS OF THE DISTANCE ASSESSMENT METHODS IN DEVELOPING THEIR LANGUAGE SKILLS ACCORDING TO SPECIALIZATION

Language Skill	Category	No.	Mean-scores	standard deviations
Reading	English	130	2.60	.471
	Arabic	130	2.40	.384
	other	130	1.56	.634
	Total	390	2.19	.676
Listining	English	130	2.55	.560

	Arabic	130	2.53	.444
	other	130	1.69	.602
	Total	390	2.26	.672
	English	130	2.60	.514
Speaking	Arabic	130	2.56	.442
	other	130	1.67	.645
	Total	390	2.28	.690
	English	130	2.58	.528
Writing	Arabic	130	2.46	.446
	other	130	1.55	.657
	Total	390	2.20	.717
The total degree	English	130	2.58	.477
	Arabic	130	2.49	.368
	Other	130	1.62	.603
	Total	390	2.23	.655

Table (11) shows apparent variation in the mean scores and standard deviations of the effectiveness of the distance assessment methods in developing their language skills because of the difference of specialization variable. To show the significance of statistical differences between the mean scores, ANOVA test was used as shown in Table (11)

TABLE 12 ANOVA FOR THE IMPACT OF SPECIALIZATION IN THE EFFECTIVENESS OF REMOTE EVALUATION ON THE DEVELOPMENT OF LANGUAGE SKILLS

Skill	Source	Total of squares	fd	Means of squares	F value	Statistical Differences
Reading	Intergroup Within Groups Total	78.271 99.500 177.771	2 387 389	39.135 .257	152.216	.000
Listening	Intergroup Within group Total	62.979 112.664 175.644	2 387 389	31.490 .291	108.166	.000
Speaking	Intergroup within group total	72.124 112.929 185.052	2 387 389	36.062 .292	123.582	.000
Writing	Intergroup within group total	82.410 117.307	2 387	41.205 .303	135.936	.000

		199.717	389			
The total degree	Intergroup within group total	73.480 93.654 167.134	2 387 389	36.740 .242	151.817	.000

Table (12) shows that there are statistically significant differences at (α = 0.05) due to the specialization in all skills and instrument as a whole. To show the conjugate statistically significant differences between the means scores, the post-comparisons were used in oral way as shown in table (13).

TABLE (13): SCHEFF'E POST- COMPARISONS FOR THE IMPACT OF SPECIALIZATION IN THE EFFECTIVENESS OF REMOTE EVALUATION ON THE DEVELOPMENT OF LANGUAGE SKILLS

Skill	Specializations	English	Arabic	Means scores
Reading	English			2.60
	Arabic	.194*		2.40
	Other	1.032*	.838*	1.56
Listening	English			2.55
	Arabic	.018		2.53
	Other	.862*	.843*	1.69
Speaking	English			2.60
	Arabic	.042		2.56
	Other	.932*	.891*	1.67
Writing	English			2.58
	Arabic	.115		2.46
	Other	1.028*	.912*	1.55
The total degree	English			2.58
	Arabic	.092		2.49
	Other	.963*	.871*	1.62

*   Significant at (α = 0.05)

Table 13 shows the following:

There are statistically significant differences at (α = 0.05) between English and Arabic in reading skill. Differences are in favor of English. Also, there are statistically significant differences at (α = 0.05) between other disciplines on the one hand and both English and Arabic on the other hand. Differences were in favor of English and Arabic in reading skill.

There are statistically significant differences at (α = 0.05) between other disciplines on the one hand and both English and Arabic on the other hand. Differences were in favor of English and Arabic in listening, speaking, writing and the total degree.

This may be due to inputs, processes and outputs. The process is the study plan which include 10 English and Arabic courses that emphasize the need to develop language skills (reading, writing, listening, speaking). These courses are: methods of scientific research in language and literature, art of writing and expression, taste of literary text, technical writing in English, Arabic Library, entrance to Arabic linguistics and graphic disgrace in Qur'an and Sunnis, teaching skills in English and selected literary texts in English and English for specific purposes.

The processes consist of teaching and evaluating students remotely but the outputs are the results which revealed differences in favor of Arabic and English as compared to non-linguistic disciplines.

The difference between Arabic and English at the level of reading skill may be attributed to the fact that four of the 10 courses are taught in English that may

not be taught by Arabic students. Also, the fact that Arabic courses are limited to the linguistic competencies of the grammatical, syntactic, morphological, phonetic, and rhetorical and semantic systems and their lack of courses that develop the linguistic performances of skills.

While the compulsory requirements of the English language plan included skills that developed skills (Reading, vocabulary building, oral skills, speaking, paragraph writing, essay writing and research writing in English. For supporting specialization requirements, there are two courses (English teaching methods and computer assisted English learning). These two courses may have a significant impact between Arabic and English in reading skill in favor of English. Besides, there are difference in means scores in all skills in favor of English so the relationship between language skills is interactive.

RECOMMENDATIONS

Design Arabic language plan that includes courses centered on language skills emulating the English language plan.

Achieving the balance between the four language skills in the English plan.

Designing questions that measure listening and speaking skills.

Designing Question that measure the advanced levels of reading and writing skills

Assigning (50%) of marks to semester work for courses taught remotely.

REFERENCES

[1]   Abu Nimreh, M.K. & Al Hadidi, M.A. (2009). The Training Needs for the Teaching Staff in the Faculties of Educational Sciences in the Private Universities in Jordan. Studies of educational Sciences, (36), 362-387.

[2]   Aljabouri, H. (2008). Evaluating Members' Stuff in the light of Professional Qualities and Personal Characteristics. UOB Journal. 15, 2, 636- 660

[3]   Aljado', F. (1999). The Phenomenon of Failure in University Education in Kingdome of Saudi Arabia. Unpublished Master Thesis. King Saud University.

[4]   Almenie, O. (2018). Academic cheating in E-learning from the point of View of Faculty of members at King Saud University. Journal of Education Studies. 16, 121-179.

[5]   Almeqren, N. (2019). The Effect of E-Learning by Using The Learning Management System Edmodo on The Achievement of The Students of first Secondary Grade In The Curriculum of Computer and Information Technology and their attitudes towards Technology. International Interdisciplinary Journal of Education. 8, 118-135.

[6]   Alothman, A. (2020). The E-Evaluation and Measurement tools: New Education. https://cutt.us/az8Z6

[7]   Alshaikhi, H. (2015). The Role of the University Professor in Improving the Quality of the Ways and Methods of Evaluating Students. Journal of Arabs Union for Education and Psychology. 13, 1, 55-89.

[8]   Bani-Yasain, M. (2018). Al - Balqa Applied University Students' Attitudes towards Their Evaluation Methods. Shaqra University Journal. 10, 1, 117-135.

[9]   Bani-Yasain, M. (2016). The aim of this study is to analyze the General Secondary examination (GSE) questions of Arabic language subject in Jordan according to Bloom's cognitive. Journal of Linguistics and Literary Studies. 7, 1, 75-92.

[10] Bani-Yasain, M. & Alzu'bi, M. (2013). The Effectiveness of Evaluation Style in Improving Lingual Knowledge for Ajloun College Students. Journal of Education and Practice. 12, 14, 73-81.

[11] Basyouni, R. (2016). The Effectiveness of Some Electronic Evaluation Tools in the Development of the Performance of Computer Teachers in the Preparatory Stage. Arabic Journal for Sciences and Research Publication. 2, 7, 55-92.

[12] Burqa'an A. & Alrabea' A. (2003). Evaluating the instructional practices of educational vocabularies from students' point of view. Hadhramout for studies and researches. 3, 4, 110.

[13] Megdadi, M. (2020). Perceptions of High School Students in Government Schools in Jordan to Use Distance Education in Light of the Corona Crisis and its Developments. Arab Journal for Scientific Publishing (AJSP). 19, 96-114.

[14] Suleiman, K. (2010). The practices University of Tabouk members in evaluating their students' achievement in the light of some variables. Journal of Umm Al-Qura University of

Educational Psychological Silences. (2)2, 353-417.