

Received on (04-09-2019) Accepted on (27-11-2019)

The Effect of Using Authentic Assessment on Eighth Grade Student's Motivation in English Language Course in Amman

Main Researcher : Mo'en salman Alnasraween

Researcher : Rami Ibrahim Shogran

Researcher: Ala Ahmed Harareshi

Researcher: Mai saleh Qattash

University Name & City

Educational psychology - - educational and psychology college

* Corresponding author:

E-mail address: mueen@aau.edu.jo

Abstract:

This study aimed to investigate the effectiveness of using authentic assessment on eighth grade students' motivation in English language course. The study sample consisted of four group, total of (200) students of the eighth grade for the second semester of academic year (2018-2019) they were divided randomly to control and experimental group. To implement this study, the researchers prepared a training program to train the teachers on the authentic assessment tools, also two tools for authentic assessment (Checklist and Rubric) were developed, also to measure the motivation, the Lipper scale was used which. The results showed statistically significant differences in favor of the experimental group in motivation due to the use of authentic assessment tools, Moreover, the results showed statistically significant differences in motivation for males in the experimental group.

Keywords: Authentic Assessment, Motivation, English Language Course.

أثر فعالية استخدام التقويم الواقعي على دافعية طلبة الصف الثامن الاساسي في مبحث اللغة الإنجليزية

الملخص

هدفت هذه الدراسة إلى استكشاف أثر فعالية استخدام التقويم الواقعي على دافعية طلبة الصف الثامن الاساسي في مبحث اللغة الإنجليزية، تكونت عينة الدراسة من أربعة مجموعات بواقع (200) طالباً من الصف الثامن للفصل الدراسي الثاني من السنة الدراسية (2018-2019) تم تقسيمهم عشوائياً لمجموعتين ضابطة وتجريبية. ولتنفيذ هذه الدراسة، أعد الباحثون برنامجاً لتدريب المعلمين على أدوات التقويم الواقعي، كما تم تطوير أداتين للتقييم الواقعي (قائمه شطب، وسلم تقدير) ، ولقياس مستوى الدافعية ، تم استخدام مقياس ليبير. وأظهرت النتائج وجود فروق ذات دلالة إحصائية لصالح المجموعة التجريبية تعزى لاستخدام التقويم الواقعي، كما أظهرت النتائج وجود فروق دالة احصائياً في مستوى الدافعية تعزى للطلبة الذكور في المجموعة التجريبية.

كلمات مفتاحية: التقويم الواقعي، الدافعية، مبحث اللغة الإنجليزية.

Introduction:

The field of measurement and evaluation is one of the most important areas of education and psychology and the most important pillars of the educational process, which aims to develop and improve the educational process in all its aspects. Measurement and evaluation include procedures, materials and tools that help in reaching comprehensive results characterized by validity and reliability and participate in making decisions related to the educational system. The importance of the evaluation process in its role in guiding the educational process and the need to make the right decisions at the level of the individual and society, as the student is the focus of the educational process and therefore the measurement and evaluation focuses on the most attention to the learning outcomes of the learner, and with continuous development has expanded to include all related to learning such as curriculum, teaching methods, and supervisory and instructional programs to include activities related to the educational process such as school broadcasting, parent councils and other programs with direct and indirect impact on students (Reid and Fitzgerald, 2011).

The evaluation has been based on traditional tests in all its forms as a means to highlight individual differences and encourage competition among students to obtain superior positions among their peers without showing the individual's actual abilities and capabilities. The main aim of the students was to move from one stage to the next. Excellence was a way to attract and gain the attention of teachers and parents and perhaps reached a reason to boast among colleagues of no real benefit in practical life (Caudle, 2016).

As a result of the failure of the traditional tests of providing feedback on the level of performance of students in order to make improvements to the educational learning process, and does not provide the information needed by the teacher to compare the performance of students according to performance standards, they have faced a lot of criticism and claims for alternative, in addition to the various pressures that the development of educational ideas and the appearance of new educational ideas, moreover the technological explosion and the rise of the information age, all that has helped in the rapid development of the field of measurement and evaluation during the past five decades thus has shown new innovations and trends in the field of measurement and evaluation in all areas, and increased emphasis on the procedures of measurement and evaluation, which assist in the development of the educational institution all its services and make them more effective, in addition to the emergence of new tools, procedures and methods of measurement that have had a positive impact on the learning process (Crocker, 2013) .

The new educational trends have led to the emergence of many new terms that describe new types of evaluation, such as authentic assessment, alternative assessment, and some called performance assessment. This assessment reflects student performance and measures it in real situations where students engage in tasks of value and meaning. Their skills through self-learning and direct interaction with the environment through different ways and means, such as individual and collective classroom projects, observation, achievement files, etc. (Cano & Ion, 2016).

Mueller (2005) asserts that evaluation in all its forms strongly influences the behavior of individuals and can be a motive for change of behavior. Therefore, the shift or change to an alternative assessment makes the educational system more coherent and consistent with other

systems especially economic and social, and this evaluation should push the trainers to focus more on linking educational content to practical life and setting academic endeavors in the context of the real world and making them meaningful for students.

The shift towards authentic assessment practice can address the negative effects of traditional evaluation; psychological and financial stress tests for students and parents; moreover, the role of students in using traditional evaluation tools is negative, while authentic assessment stimulates students' ability to produce ideas and solve problems. Authentic assessment as a term is used to describe real tasks that require students to perform and / or produce knowledge rather than reproduce information discovered by others. This assessment also focuses on meaningful tasks, positive interaction, knowledge integration, Self-reflection and real-life learning. Authentic assessment can be used to plan and implement creative and authentic learning at the highest levels of learning. Authentic assessment is an ideal complement to education that meets the needs of all students at different levels of the classroom (Hildebrand, 2005).

The creation of authentic assessment concept is credited to Grant Wiggins (1989), which he defines as any assessment task that uses multiple assessment systems to measure students' behavior and memoir in meaningful tasks related to extracurricular life (Gulikers, Bastiaens, & Kirschner 2004). Authentic assessment according to Wiggins (1993) leads students to meet the challenges of future professions and to analyze and solve life problems in a professional manner. Authentic assessment enables teachers to learn about students' weaknesses and to address problems which they face and take meaningful actions to address and resolve those problems.

Research problem:

The purpose of this study is to identify the effect of the use of authentic assessment on the motivation of eighth graders in English language.

Research Questions:

The study will attempt to answer the following questions:

Question 1: What is the effect of using authentic assessment strategies on the motivation of eighth graders towards English language?

Question 2: Does the effect of using authentic assessment strategies differ in students' motivation towards English language by gender?

Significance of the Study:

This study has a theoretical and practical significance as follows:

Theoretical Significance:

This study may contribute to enriching the Arab library by shedding light on the nature of the authentic assessment and its impact on students' motivation and increasing interest of researchers and those interested in it. This study is an opportunity to introduce teachers to the authentic assessment role on motivation; it also shows whether there is a difference in the effect of authentic assessment of motivation between females and males.

Practical Significance:

- This study will provide a training program on two of authentic assessment tools: check list and rubric.
- This study will provide two tools for the authentic assessment of the eighth grade English language course, namely the checklist and rubric that teachers and users can use.
- Taking over the lack of motivation for learning among eighth graders in the English language course with authentic assessment.
- Applying the tools of authentic assessment of the English language of the eighth grade, which had been prepared by the researchers.
- Opening the horizons for researchers to learn about the methods and tools of authentic assessment, and conduct more studies related to the authentic assessment and motivation for learning.

Limitations and Delimitations of the Study:

- This study was limited to the basic eighth grade students in the schools of Amman Kasbah in the capital Amman Governorate during the second semester 2017/2018. The study was limited to the English language and the assessment tools prepared by the researchers for the purposes of study. The results of this study are limited to the reliability and validity of the used tools.

Terminological and Procedural definitions:

Authentic Assessment:

Gulikers, Bastiaens and Kirschner (2004, 71) have identified the authentic assessment as "an assessment that requires students to use the same competencies and integrate them with the knowledge, skills and attitudes that they need to apply in the standard setting of professional life."

Mueller (2014, 62) has defined it as "a form of assessment that requires students to perform tasks in the real world that demonstrate a meaningful implementation of basic knowledge and skills."

Wiggins (1993, 14) has defined it as "... engaging worthwhile problems or questions of importance, where students must use knowledge to formulate performance effectively and creatively, and tasks are either identical or similar copies of the type of problems faced by Adults, consumers or professionals in this area. "

The procedural definition: the use of non-traditional assessment tools, namely, check list and rubric to measure student achievement in the required subject.

Check list: Assessment tools with specific criteria that teachers and students may use to measure skills development and progress. Checklist identify the skills, attitudes, strategies and behaviors to be evaluated and provide methods for organizing information systematically and may be individual or association. In general, the check list consists of a set of data that meet specific criteria; the answer to each statement is "yes", "no", "satisfactory" or "unsatisfactory" etc. (Stiggins, 2001, 86).

The procedural definition: a set of criteria designed to measure student achievement in speaking and listening skills.

Rubric: is a recording tool that explicitly represents the performance expectations of a particular task or skill. Each paragraph is subtracted to several levels ranging from lack of proficiency to full

mastery and varying degrees of proficiency. For skill or task characteristics, a rubric can be used for a wide range of tasks: handout sheets, projects, oral presentations, technical performances, group projects, etc. The gradient models can be used to provide feedback to support or guide continuous learning efforts or both (Reddy & Andrade, 2010, 21).

The procedural definitions: a set of criteria designed to measure students' proficiency in the following English language skills (reading, writing, grammar and vocabulary, organization and task completion).

Motivation: Sternberg & Williams defines motivation as described in (Qatami, Shreem, Ghuraibah, Zubay, and Matar Waqaza, 2010, p. 287) as: "inner energy or mental power that helps the person achieve goals within different contexts, school, home, the whole world. "

In this study is defined as the degree to which the student obtains a response to the motivation scale used in this study.

Related Studies:

In a study conducted by Beni Khalid (2013) to study the effect of teaching on R  thkov's (texts) in the achievement of the tenth grade students and their motivation to learn, the researcher chose the sample of the study which consisted of (66) students by random way, the study sample was divided into two groups; The first group was conducted using R  thkov's learning strategy, while the second group was studied in the traditional way, and the researcher constructed the achievement test and also used the motivation scale for learning for Koziki and Antosil for the purpose of collecting the study data. The results of the statistical analysis that there were differences Statistical significance in both achievement and motivation for learning in favor of the experimental group which was studied using the pattern of education by means of written materials (texts) of R  thkov's.

Andrade, Wang, Du, & Akawi (2009) conducted a study to investigate the impact of the use of self-assessment based on a set of self-efficacy guidelines. The study sample consisted of (268) primary and secondary school students in the State of New York, the researchers prepared an index containing a set of guidelines containing several criteria that students must follow when evaluating their own work and prepared the self-efficacy scale. The results of the study showed that there were statistically significant differences in self-efficacy for the experimental group, and there were statistical differences in gender for females where their self-efficacy was higher.

Zimbicki (2007) examined the goal of evaluating the impact of the alternative evaluation on students' motivation and self-efficacy. The study sample consisted of (72) seventh graders. In addition to the alternative assessment tools, the researcher used a questionnaire to measure the motivation of the students, in addition to the interview. The results indicated that there were statistically significant differences in favor of the alternative assessment, which had a clear effect on students' motivation and self-efficacy.

Margulies & Ghent (2005) conducted a study to investigate the impact of alternative assessment strategies on student absorption in biology. The sample of the study was (50) male and female students. The sample of the study was divided into (25) students and experimental sample was taught using the alternative assessment method, and (25) male and female control sample were taught in traditional ways. The researchers used the observation and the file of achievement as tools for study. The study showed statistically significant differences in favor of the experimental group

in improving students' performance, increasing their academic achievement and reducing behavioral problems.

A study by Suurtamm (2004) aimed to investigate the effect of authentic assessment on the practices of mathematics teachers. The researcher tracked five teachers of secondary school who participated in a training program for the methods of applying authentic assessment in the teaching of mathematics in their classrooms in Ontario, the researcher relied on observations and interviews to gather information. Teachers have become more aware of the connection between mathematics and their ideas in the real world, and students have a responsibility to learn and develop their learning of mathematics and to increase their self-confidence.

Mintah (2003) conducted a study to investigate the prevalence of physical assessment in physical education and its impact on student motivation, achievement skills and self-concept. From the teachers' point of view, the sample consisted of (108) male teachers and (102) female teachers' Public schools in Blivard with a total of (210) of the study society, which was (396) teachers. The researcher has constructed a questionnaire to collect information. The results of the study showed that the authentic assessment is spread in the public schools in that region satisfactorily. The results also showed that the authentic assessment positively affects students' self-concept, motivation and achievement of tasks. The researcher also recommended conducting research to investigate the students' perceptions of the authentic assessment.

Mcdonald & Boud (2003) conducted a study aimed at investigating the impact of high school students' training in Sydney, Australia on the self-assessment process. The sample consisted of (512) students distributed in (10) schools, the sample of the study was divided into experimental group (256) students and the control was (256) students. The experimental group was trained in self-assessment skills, and the control remained in the traditional style. The results showed that there were statistically significant differences in favor of the students who were trained as they improved achievement level.

In a study conducted by Newman, Bryk and Nagaoka (2001), the effect of real homework and authentic assessment of students' intellectual work in their daily school performance and student achievement in standardized tests, Of the 19 primary schools in Chicago, the study included about (5,000) students and their teachers in the third, sixth and eighth grades. These classes were deliberately chosen because of the appropriateness of using test scores from each of the state-level test programs, this allowed researchers to link teachers' duties to student performance in state reading and math tests and results from standardized national tests for reading and mathematics. In addition to the test scores, the teachers' tasks were analyzed in writing and mathematics. A group of teachers from the Chicago Public Schools were trained to judge the quality of teachers' tasks using evaluation models that included intellectual work standards and found that when teachers organized real- Students not only acquired intellectual work skills, but also gained higher grades in state reading and math tests. They also noted that intellectual demands for teachers' tasks or assessment tasks played a much more important role than a particular teaching strategy or teaching method to influence and therefore the professional development of teachers should focus on their ability to design and use curriculum materials and classroom assessment that includes a real high intellectual challenge.

Through a review of previous studies, the diversity of methods and areas of research in the subject of authentic assessment by researchers is evident. Most of the studies were similar in their tools. Zambicki (2007), Margulies & Ghent (2005) Mcdonald & Boud (2003) used authentic

1090

assessment tools as an essential tool in the study. In addition to additional tools that differed according to the dependent variable, Marcioles and Ghent (2005), Suurtamm (2004), used observation and interview as tools for study, and the studies of Minta (2003), Andrade, Wang, Du, & Akawi (2009) used the self-efficacy scale, Beni Khalid (2013) and Zimbicki (2007) used a scale of motivation as well as the above tools.

The previous studies aimed to investigate the effect of authentic assessment on one or more variables related to the educational process; Beni Khalid (2013) study aimed to investigate the effect of authentic assessment on students' achievement; and Zimbicki (2007), Minta (2003), Andrade, Wang, Du, & Akawi (2009) aimed to investigate the impact of the authentic assessment of students' self-efficacy, and Zimbicki (2007), Minta (2003) and Beni Khalid (2013) on the impact of authentic assessment on motivation. In addition to the above, the Newman, Bryk, Nagaoka (2001) aimed to assess the effect of authentic homework and authentic assessment on students' intellectual work in their daily school performance and student achievement in standardized tests.

Methodology of the study:

This study used a semi-experimental approach to achieve the purpose of this study.

Study population:

The population of the study consisted of eighth-grade students in the district of capital Amman in Aljameha directorate of education. Which according to the latest statistical report of the Ministry of Education (3883) male and female; 1778 males and 2105 females distributed in (71) schools.

The following table illustrates the division:

Table (1): Distribution of the Study Society

	Males	Females	Total
Number of students	1778	2105	3883
Number of schools	37	34	71

2. Study Sample:

The sample of the study consisted of four groups randomly selected with (200) students (5%) of the study population for the second semester of the academic year (2017-2018) distributed as follows: (2) female groups were randomly divided into control group and experimental sample, and (2) male students group randomly divided into a control group and an experimental group.

Table (2): Distribution of the sample of the study by gender and group

Gender Sample	Male	Female	Total
experimental	50	50	100
control	50	50	100
Total	100	100	200

Study Tools:

Training Program:

The researchers prepared a training program to train teachers on authentic assessment tools. The program included an introduction to measurement and evaluation, the categories of educational measurement and its purposes, the concept of educational evaluation and its characteristics, authentic assessment and its benefits and strategies, authentic assessment tools, in addition to a detailed explanation of the Rubric and Check list and the mechanism of trade-off between them. The program included examples and practical exercises. The duration of the program is (10) training hours; was divided into two training days, (5) hours of training per day, the goal of this program is to reach the trainees to a good degree of knowledge of authentic assessment methods that enable them to transfer that knowledge and applied in the classroom.

The validity of the training program:

In order to verify the validity of the training program, the researchers presented it to a group of faculty members at Jordanian universities, also it was presented to some educational supervisors of the English language at the Ministry of Education and UNRWA to ensure the comprehensiveness of the training program and its relevance to the subject, The adjustment were made in the light of the observations made.

Motivation scale:

The researcher used scale of motivation that developed to the Jordanian environment by Alwan and Al-Attiyyat (2010); the scale was developed by (Lepper, 2005), which consisted of three major domains: Preference of challenge, Curiosity, and a desire for independent mastery. The scale consists of (24); (8) items per domain, and the student answers the scale through a five-point scale: strongly agree , agree, unsure, disagree, highly disagree, and the corresponding scores for these alternatives are: (5,4,3,2,1), respectively, and the grade for each domain is between (8) the lowest degree, and (40) the highest degree.

Content Validation:

The scale was presented in its preliminary form to a group of specialized arbitrators in education, psychology and measurement in some Jordanian universities to ascertain the veracity of the measurement aspects of the motivation of learning and its applicability in determining the level of motivation. Comments agreed upon by 80% of the arbitrators were amended, and the scale was then prepared in final form.

Reliability of the scale:

The reliability of the scale has been verified in two ways:

- Test-retest method: where the scale of motivation was applied on a sample survey consists of (50) students from the study population and outside the sample then the scale was applied on them and after two weeks it was reapplied on them again and then Pearson correlation between implementation was calculated.
- Reliability of internal consistency: the internal consistency of the study scale was confirmed using the Cronbach Alpha coefficient, and Table (4) shows the values of these consistency coefficients.

Table (4): Values of correlation coefficients and internal consistency of the motivation scale

#	Domain	Pearson correlation	Internal consistency
1	Preference of challenge	0.80	0.78
2	desire for independent mastery	0.82	0.87

3	Curiosity	0.78	0.80
Total Score		0.89	

The results of Table (4) show that Pearson correlation coefficients were 0.89 for the total score and for the domains ranged between 0.82 and 0.78. The internal consistency values ranged between 0.87 and 0.78. All of which are appropriate values for the purposes of the present study.

Authentic Assessment Tools:

Two tools were used from authentic assessment tools (Rubric and Check list). The researchers built the assessment tools to measure the achievement of students in the fifth module of the eighth grade English course (You Can Do It). They studied the theoretical literature concerned with the methods of building the authentic assessment tools, and then analyzed the content for the required unit of study and made a list of educational objectives for the required material using the curriculum book and the teacher's guide and then drafting the required standards for building the two tools. To verify the validity and reliability of the tools, the researchers presented the two tools to a group of English language teachers in the schools of the Ministry of Education and UNRWA, the educational supervisors of the same subject, and some professors in the Jordanian universities to take their opinions and observations. Comments agreed upon by 80% of the arbitrators were amended, and then the researchers prepared the tools in the final version as follows:

1. **Rubric:** The scale is of 5 main criteria that measure the following skills: Reading, Writing, Grammar, Punctuation, Vocabulary and Spelling, Organization, and Task. Standard to (4) levels ranging from lack of proficiency to mastering is given to each grade level and then the total score is calculated to extract the final score.
2. **Checklist:** It consists of only two criteria that measure listening and speaking skills. Each criterion is divided into 4 levels ranging from non-proficiency to proficiency and is given to each grade level and then the total score is calculated to extract the final score.

For the purposes of statistical analysis, skills in the tools were grouped into four basic domains:

1. Reading and Vocabulary.
2. Listening.
3. Writing and Organization.
4. Rules and task completion (Grammar and Task Completing).

Study Procedures:

The researchers followed the following procedures:

1. Review the theoretical literature, studies and measurements related to the variables of the study; authentic assessment, motivation.
2. Build the training program and ensure its validity and reliability.
3. Building the tools of the study and ensure the validity and reliability.
4. Training teachers on the evaluation tools used through the program prepared for this purpose.

5. Selection of the study sample from the study community and distribution of the measure of motivation to the sample after explaining the purpose and purpose of the study and how to answer it.
6. Agree with teachers on the plan to implement custom assessment strategies (authentic or traditional) and implementation mechanism.
7. Apply authentic assessment strategies and tools to the experimental sample and obtain and review the results and verify their validity.
8. The distribution of the motivation scale on the sample for the second time to compare the results.

Study Results:

Results of answering the first question of the study: to answer this question, the arithmetic means and the standard deviations of the initial and post implementation of the motivation of the eighth-grade students towards the English subject were extracted according for the two groups of study.

Table (5): The arithmetical means and standard deviations of the students' performance of the initial and post implementation on the motivation scale

group		Pre-test		Post-test	
		mean	Standard Deviation	mean	Standard Deviation
Preference of challenge	Control	3.68	0.84	3.89	0.77
	Experimental	3.45	0.77	4.07	0.64
Curiosity	Control	3.80	0.72	3.97	0.75
	Experimental	3.63	0.75	4.20	0.64
desire for independent mastery	Control	3.80	0.70	3.97	0.67
	Experimental	3.65	0.71	4.10	0.65
Total score	Control	3.77	0.69	3.95	0.67
	Experimental	3.60	0.69	4.14	0.58

The results of Table (5) show that there are observed differences between the arithmetic mean of the motivation of the groups. The arithmetic means of the experimental group reached 4.14 and for the control group reached 3.95. To test the significant of those differences, the Multivariate Analysis of covariance was extracted and Table (6) shows this.

Table (6): Results of (MANCOVA) to examine the significance of the differences between the arithmetical means of the students' performance on the Post implementation of the motivation scale according to the group

		Sum of Squares	df	Mean Squares	F	Sig.	Eta Squared
Initial implementation	Preference of challenge	0.445	1	0.445	0.891	0.346	0.005
	Curiosity	1.641	1	1.641	3.395	0.067	0.017
	desire for independent mastery	0.721	1	0.721	1.671	0.198	0.008

	Total score	0.623	1	0.623	1.583	0.210	0.008
Group	Preference of challenge	1.816	1	1.816	3.640	0.058	0.018
	Curiosity	3.252	1	3.252	6.729	0.010	0.033
	desire for independent mastery	0.971	1	0.971	2.250	0.135	0.165
	Total score	2.063	1	2.063	5.245	0.023	0.189
Error	Preference of challenge	98.302	197	0.499			
	Curiosity	95.219	197	0.483			
	desire for independent mastery	85.016	197	0.432			
	Total score	77.509	197	0.393			
Total	Preference of challenge	3270.438	200				
	Curiosity	3436.016	200				
	desire for independent mastery	3344.781	200				
	Total score	3355.384	200				
Corrected total	Preference of challenge	100.367	199				
	Curiosity	99.592	199				
	desire for independent mastery	86.518	199				
	Total score	79.944	199				

The results of Table (6) indicate that the value of (F) for the total score of the group reached (5.245) which is a statistical significant at ($\alpha = 0.05$), And for the domain "curiosity", the value of (F) is (6.729) which is a statistical function at (0.05), also the values of the ETA squared ranged between (0.005) and (0.033) which indicate a moderate effect of the program on the student motivation. In order to determine the variance of the differences for the group, the modified arithmetic means, and standard errors were extracted as shown in Table (7).

Table (7): Modified arithmetic means and standard errors of students' performance on the post-implementation of the motivation scale

		Means	Standard Errors
Curiosity	control group	3.96	0.070
	Experimental group	4.21	0.070

Total Score	control group	3.94	0.063
	Experimental group	4.15	0.063

The results of Table (7) show that the mathematical means of the post- implementation of the experimental groups are higher than the control group for the total score and the "curiosity" domain.

First: Discussion the results of the first question:

The results of the answer to this question showed that there were statistically significant differences in student motivation for the experimental group. This result is due to the use of authentic assessment strategies.

The researchers argue that the increased motivation of the experimental group, which has been studied using authentic assessment strategies, may be due to the role of authentic assessment in allowing students to participate and integrate into the educational process as the student is the focus of the educational process. It is clear to the student that the authentic assessment strategies allow students to make errors and learning through these errors for their future, So the student becomes more interested and eager to learn more, and the authentic assessment also allows for teamwork, exchange of experiences, and alleviation of anxiety among students due to lack of specificity of time, joint work and freedom in choosing the type of evaluation. Burke (2009) noted that authentic assessment focuses on output as well as quality of performance, and students are more actively involved in the learning process. In addition, students know how they will be assessed before the actual assessment, and thus higher levels of motivation and attention. This result agrees with the results of the Beni Khalid study (2013), which aimed to study the effect of teaching by the written materials (texts) of Röthkov. on the 10th grade and their motivation to learn. The results showed statistically significant differences in achievement and motivation for learning for the experimental group Studied using the teaching style by the written materials (texts) of Röthkov. It also agreed with Zimbicki (2007), which aimed to test the effect of the alternative assessment on students' motivation and self-efficacy, the results indicated that there were statistically significant differences in favor of the alternative assessment, which had a clear effect on student motivation and self-efficacy. It also agreed with the Hancock study (2007), which aimed to assess the performance of motivation and achievement of postgraduates. The results revealed that students who had been evaluated performed somewhat higher in the final exam and showed higher levels of Motivation to learn from students who have been evaluated through traditional tests. It also agreed with the Minta study (2003), which aimed to investigate the prevalence of physical assessment in physical education and its impact on students' motivation, achievement skills and self-concept from the point of view of teachers, the results showed that the authentic evaluation positively affects the self-concept of students and their motivation and achievement of tasks. It also agreed with Marra & Wheeler (2000), which used authentic assessment / projects and traditional teaching methods for two full years, and aimed to maintain a common knowledge base on the scientific / engineering message among students and to allow students to focus on Knowledge and skills of interest to them, and the results showed a high level of motivation and enthusiasm for the project of students in the experimental group that adopted the project method.

Results of the second question: To answer this question, the arithmetical means and standard deviations were extracted, and Table (8) shows this.

Table (8): means and standard deviations of student performance on the motivation scale by group and gender

group			Pre-Test		Post -test		Number
			mean	deviation	mean	deviation	
desire for independent mastery	control group	Male	3.65	.890	3.87	0.80	50
		Female	3.71	.810	3.92	0.74	50
		Total	3.68	.840	3.89	0.77	100
	experimental group	Male	3.35	.770	4.30	0.50	50
		Female	3.56	.770	3.84	0.69	50
		Total	3.45	.770	4.07	0.64	100
Curiosity	control group	Male	3.74	.710	3.82	0.86	50
		Female	3.87	.730	4.11	0.61	50
		Total	3.80	.720	3.97	0.75	100
	experimental group	Male	3.44	.790	4.30	0.52	50
		Female	3.82	.670	4.11	0.74	50
		Total	3.63	.750	4.20	0.64	100
desire for independent mastery	control group	Male	3.74	.760	3.86	0.79	50
		Female	3.85	.650	4.09	0.51	50
		Total	3.80	.700	3.97	.670	100
	experimental group	Male	3.49	0.77	4.30	.480	50
		Female	3.81	0.60	3.90	.730	50
		Total	3.65	0.71	4.10	.650	100
Total Score	control group	Male	3.72	0.71	3.86	.780	50
		Female	3.82	0.67	4.05	.530	50
		Total	3.77	0.69	3.95	.670	100
	experimental group	Male	3.43	0.73	4.32	.450	50
		Female	3.76	0.60	3.97	.650	50
		Total	3.60	.690	4.14	.580	100

The results of Table (8) indicate that there are apparent differences between the arithmetic mean in the total score of the Post implementation of the motivation scale of the group and gender. The means for females in the experimental group (3.97) and for the males in the experimental group (4.32), to find out if these differences are statistically significant, the analysis of multiple variance has been extracted and table (9) shows that.

Table 9: Analysis of (MANCOVA) for the differences between the mean scores of the two study groups on the post-implementation of the motivation scale by group and gender

		Sum of Squares	df	Mean Squares	F	Sig.	Eta Squared
Post implementation	Preference of challenge	1.122	1	1.122	2.369	.1250	
	Curiosity	1.929	1	1.929	4.093	.0440	
	desire for independent	1.306	1	1.306	3.219	.0740	

	mastery						
	Total	1.113	1	1.113	2.973	.0860	
Group	Preference of challenge	1.960	1	1.960	4.139	.0430	
	Curiosity	3.311	1	3.311	7.026	.0090	
	desire for independent mastery	1.051	1	1.051	2.590	.1090	
	Total	2.168	1	2.168	5.789	.0170	
Gender	Preference of challenge	2.503	1	2.503	5.287	.0230	.0260
	Curiosity	.0140	1	.0140	.0300	.8620	.0000
	desire for independent mastery	.5740	1	.5740	1.415	.2360	.0070
	Total	.5490	1	.5490	1.466	.2270	.0070
Group * Gender	Preference of challenge	3.553	1	3.553	7.505	.0070	.0370
	Curiosity	3.296	1	3.296	6.994	.0090	.0350
	desire for independent mastery	5.358	1	5.358	13.204	.0000	.0630
	Total	3.973	1	3.973	10.610	.0010	.0520
Error	Preference of challenge	92.324	195	.4730			
	Curiosity	91.902	195	.4710			
	desire for independent mastery	79.129	195	.4060			
	Total	73.025	195	.3740			
Total	Preference of challenge	3270.438	200				
	Curiosity	3436.016	200				
	desire for independent mastery	3344.781	200				
	Total	3355.384	200				
Corrected Total	Preference of challenge	100.367	199				
	Curiosity	99.592	199				
	desire for independent mastery	86.518	199				
	Total	79.944	199				

The results of Table (9) indicate that the value of (F) for the gender was (1.466) and is not statistically significant at ($\alpha = 0.05$). The value of (F) is not statistically significant for all domains except for "Preference of challenge", to find out about the differences in this domain, the modified arithmetical means and standard errors were extracted as shown in Table (10).

Table (10): arithmetic adjusted means and standard errors for the performance of students on the motivation scale

		Adjusted mean	Standard errors
Preference of challenge	Males	4.095	.0690
	Females	3.868	.0690

Notes from the results table (10) that the differences between the arithmetic means on "preference challenge" domain due to the males as arithmetic means for males (4.095) and females (3,868).

For the study of the interaction between the group and the gender on the post implementation of the motivation scale, the following graphic representations were obtained:

First: the "preference challenge" domain

Figure (1): Graphical representation of the differences between the arithmetical means of students' performance at the "challenge preference" domain by group and gender.

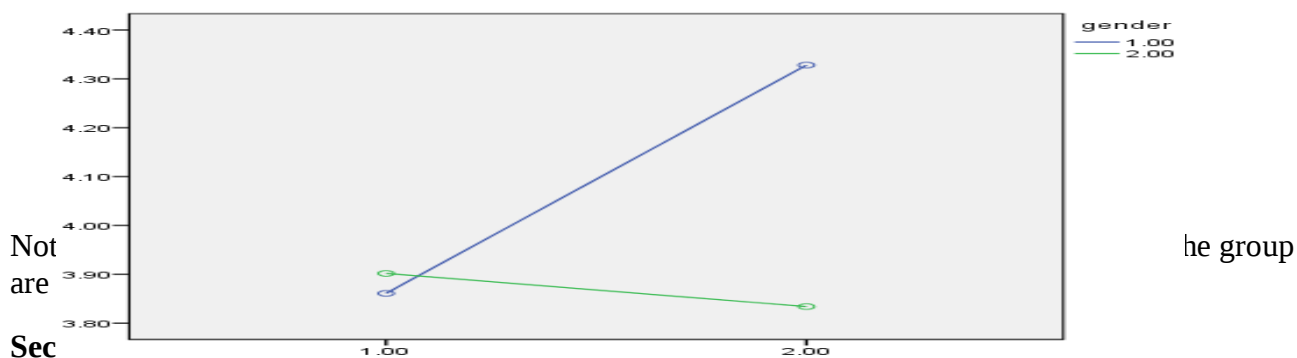
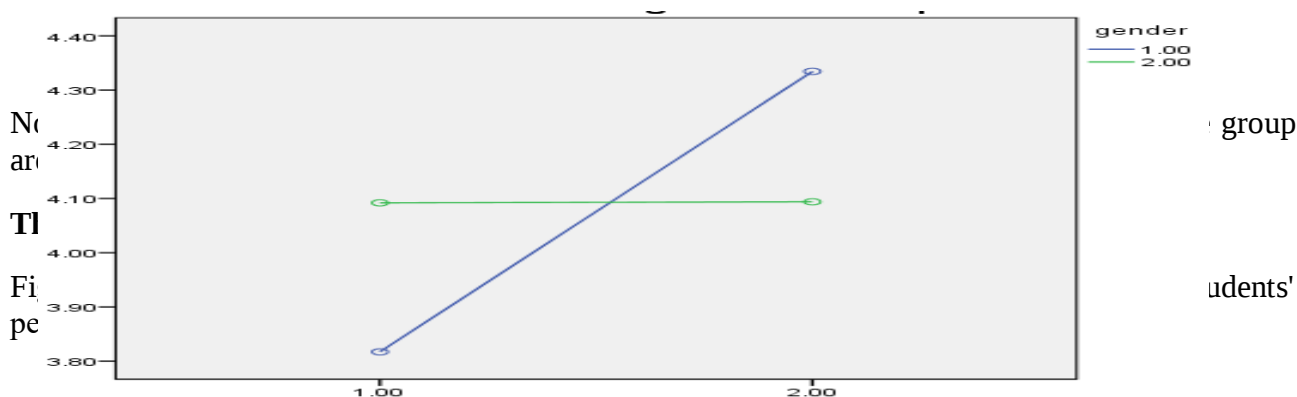
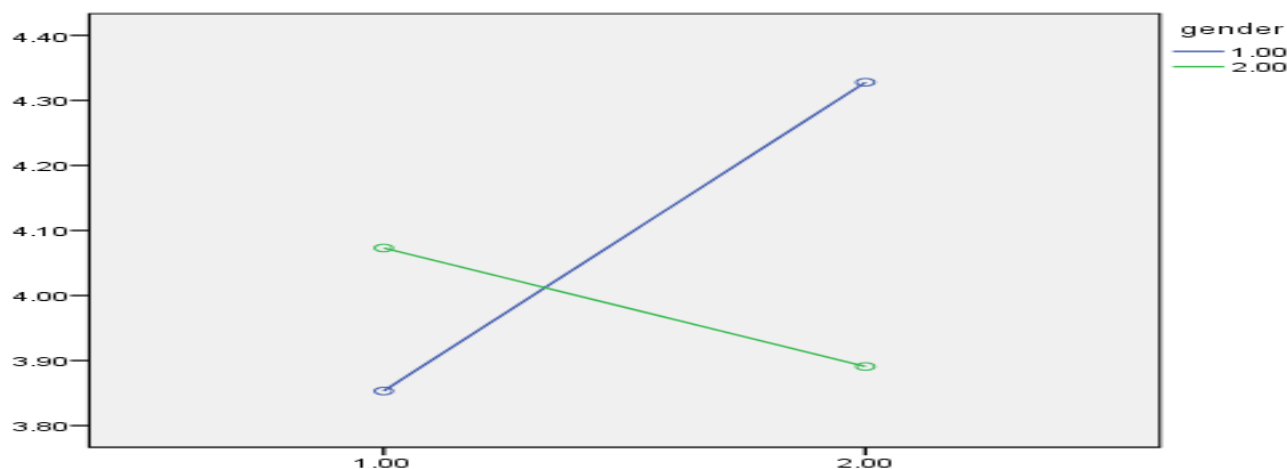


Figure (2): Graphical representation of the differences between the arithmetical means of the performance of the students at the "curiosity" domain by group and gender.

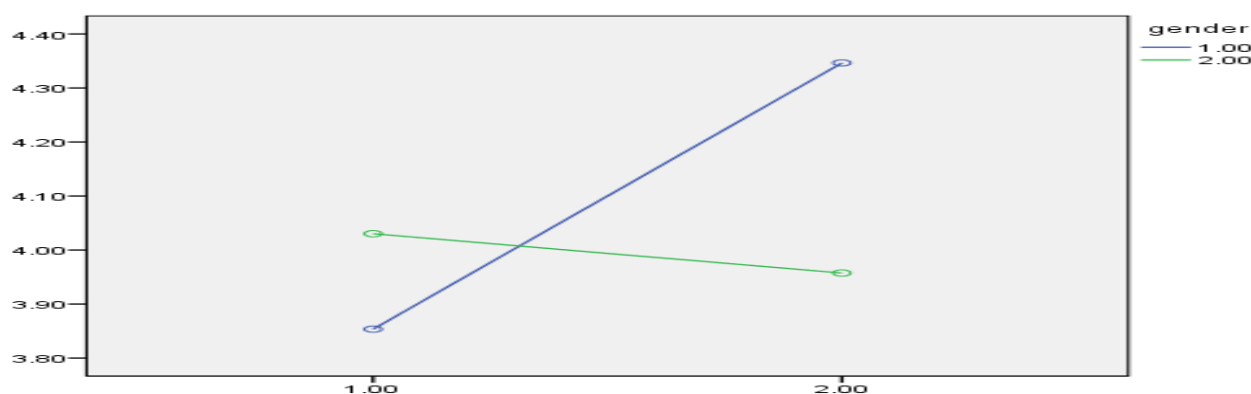




p

Fourth: The total score of motivation

Figure 4: Graphic representation of the differences between the arithmetic means of the students' performance on the total score of motivation scale depending on the group and gender.



ie

nt

differences in the arithmetic means in the motivation of students to learn English for males. This result may be attributed to the use of authentic assessment strategies.

In the view of the researchers that the increase in motivation in the experimental group in favor of males may be because of the preference of males to new challenges and unfamiliar, especially since the authentic assessment is similar to the real work in professional contexts and also highlights the situational and contextual knowledge, including the acquisition of positions and professional related competencies, it is also performance-based and requires students to demonstrate proficiency in professional practices and integrates the skills required in practical life with academic requirements, and emphasizes the evaluation for learning purposes and not for grades, and integrates the processes of social, cognitive and contemplative learning. However, this result differed with the results of most previous studies indicating that there were no statistically significant differences between females and males. A study by Amedi (2013) aimed at analyzing the motivational attitudes of learners towards English and the potential relationship between gender and motivation, the study showed that when it comes to reading, listening and speaking English, there are no significant differences between genders. Both students tend to listen, speak, read and watch English television. The overall result of the study shows that there are no differences between the genders when it comes to motivation, the results differed with other studies. Collins, Kenway, and McCloud (2000) noted that there were gender differences in students' participation in classroom work and their motivation for achievement in favor of females. On average, females outnumber

boys on several topics, this finding is consistent with the results of Martin (2004), which aimed to explore stability in factor structure, clusters, and sensory mapping in students' motivation as a gender function. The study data clearly indicated differences in motivation between males and females for females, it also agrees with the results of a study by Ghazvinia and Khajepoura (2011) to study Iranian students' attitudes and motivations in learning English. The results showed that female motivation is higher than male motivation and that males are more effective at learning English.

Recommendations:

In light of the study results, the researchers recommend:

- Raising awareness about authentic assessment strategies and tools and its positive impact on student motivation.
- Holding workshops and training programs for teachers on authentic assessment strategies and the mechanism for building tools and techniques for their use.
- Training new teachers on authentic assessment strategies before actually enrolling in school work.
- Conduct more similar studies to study the impact of authentic assessment in student motivation and achievement to all students' levels.

References:

- Alwan, Ahmed and Al-Attiyat, Khaled (2010). The relationship between academic internal motivation and academic achievement in a sample of students of the tenth grade in Ma'an in Jordan. *Journal of the Islamic University (Series of Humanities Studies)*, Volume (18), Issue (12): pp. 683-717.
- Amedi, Sandra Dokic. (2013). *The Teacher Education The C-Level of English Linguistics Elective Course: Language and Gender*. Kristianstad University. Published Thesis.
- Andrade, Heidi, Wang, Xiaolei, Du, Ying, Akawi, Robin L. (2009). Rubric-Referenced Self-Assessment and Self-Efficacy for Writing. *The Journal of Educational Research*, 102(4), PP. 287-304.
- Bani Khalid, Mohammed (2013). The effect of the pattern of education through the written materials (texts) of R  thkov. In the achievement and motivation of learning in the life sciences on students in the primary stage. *Al-Manara*, vol. 20, No. 1 (b): pp. 235-269.
- Burke, Kay. (2009). *How to Assess Authentic Learning*. (5th Ed). Thousand Oaks, CA: Corwin.
- Cano, Elena, & Ion, Georgeta. (2016) *Innovative Practices for Higher Education Assessment and Measurement*, IGI Global.
- Caudle, L. (2016). *Effective Authentic Assessment Strategies in Early Childhood and Elementary Classrooms*, Retrieved Online From: [Http: Http://Www.Wcu.Edu](http://Www.Wcu.Edu), 2/1/2018.
- Collins, C., Kenway, J., & Mcleod, J. (2000). *Factors Influencing the Educational Performance of Males and Females in School and Their Initial Destinations After Leaving School*. Canberra: DETYA.
- Crocker. (2013). Authentic Assessment: Evaluating "Real-Life" Applications of Knowledge in Higher Education, *Reflection*, 69(3), PP.44-60.
- Cumming, J., & Maxwell, G. S. (1999). Contextualizing Authentic Assessment. *Assessment in Education: Principles, Policy & Practice*, 6(2), PP. 177-194.
- Ghazvinia, Sayid & Khajepoura, Milad. (2011). Attitudes and Motivation in Learning English As Second Language in High School Students. *Procedia Social and Behavioral Sciences*, 15(2011), PP. 1209–1213.

- Gulikers, J., Bastiaens, T., & Kirschner, A. (2004). A Five-Dimensional Framework for Authentic Assessment. *Educational Technology Research and Development*, 52(3), PP. 67–86.
- Hancock Dawson R. (2007). Effects of performance assessment on the achievement and motivation of graduate students. *Active Learning in Higher Education*. Volume: 8 issues: 3, page(s): PP. 219-231.
- Hildebrand, G. (2005). Aligning Enacted Principles with Advocated Principles: Authentic and Peer Assessment in Teacher Education. Paper Presented at *The First International Conference on Enhancing Teaching and Learning Through Assessment*, Hong Kong Polytechnic University, Hong Kong.
- Luke, A., Freebody, P., Lau, S., & Gopinathan, S. (2005). Towards Research-Based Innovation and Reform: Singapore Schooling in Transition. *Asia Pacific Journal of Education*, 25(1), PP. 5–28.
- Margulies, B. & Ghent, C.A. (2005). Alternative Assessment Strategy and Its Impact on Student Comprehension in an Undergraduate Microbiology Course. *Microbiology Education*. 6, PP. 3-7.
- Marra, Rose & Wheeler, Tim. (2000). The Impact of An Authentic, Student-Centered Engineering Project on Student Motivation. 30th ASEE/IEEE Frontiers in Education Conference. F2C-13.
- Mcdonald, B. & Boud. D. (2003). The Impact of Self-Assessment on Achievement: The Effect of Self-Assessment Training on Performance in External Examinations, *Assessment in Education: Principles, Policy & Practices*, 10(2), PP. 209-221.
- Mintah, Joseph. (2003). Authentic Assessment in Physical Education: Prevalence of Use and Perceived Impact on Students' Self-Concept, Motivation, and Skill Achievement, *Measurement in Physical Education and Exercise Science*, 7(3), PP. 161-174.
- Mueller, J. (2005). The Authentic Assessment Toolbox: Enhancing Student Learning Through Online Faculty Development, *Journal of On-Line Learning and Education*, 1(1), PP. 13-26.
- Mueller, J. (2014). *Authentic Assessment Toolbox*. North Central College, Naperville, IL.
- Newmann, F., Bryk, S., & Nagaoka, K. (2001). *Authentic Intellectual Work and Standardized Tests: Conflict or Coexistence? Improving Chicago's Schools*. Chicago: Consortium on Chicago School Research.
- Qatami, Youssef and Qatami, Naifah and Shreim, Raghda and Ghraibeh, Ayesh and Zoubi, Rafta and Mattar, Jihan and Zaza, Haidar (2010). *Educational Psychology*, Jordan: Dar Wael.
- Reddy, Y., & Andrade, H. (2010). A Review of Rubric Use in Higher Education. *Assessment & Evaluation in Higher Education*, 35(4), PP. 435-448.
- Reid, C. & Fitzgerald, P. (2011). *Assessment and Employability*. London: Higher Education Academy.
- Stiggins, R. (2001). *Student-Involved Classroom Assessment*. (3rd Ed.) Upper Saddle River, NJ: Prentice-Hall.
- Suurtamm, C. (2004). Developing Authentic Assessment: Case Studies of Secondary School Mathematics Teachers' Experiences. *Canadian Journal of Science, Mathematics and Technology*, 4 (4): 497-513.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70(9), 703–713.
- Wiggins, G. (1993). *Assessing Student Performance*. San Francisco: Jossey-Bass Publishers. P. 229.
- Zimbicki, Dana (2007). *Examining the Effects of Alternative Assessment on Student Motivation and Self-Efficacy*. Unpublished Dissertation, Walden University.