

Original Article

Disempowerment Role of Assessment in English Section of Iranian University Entrance Exam: EFL University Students Perspective in Focus

Mahbubeh Mahmoodi-Nasrabadi¹, Omid Tabatabaei^{1,}, Hadi Salehi¹, Mahmud Mehrabi¹*

¹English Department, Najafabad Branch, Islamic Azad University, Najafabad, Iran

Submission date: 22 October, 2020

Acceptance date: 6 February, 2021

Abstract

The University Entrance Exam (UEE) in Iran, a high-stakes test in a multiple-choice form, has a significant effect on its stakeholders. This mixed-methods study investigates how freshman TEFL pupils experienced this nation-wide type of assessment and its (dis)empowerment role-taking advantage of Messick's framework. One-hundred freshmen university TEFL students as participants of this study were conveniently and purposefully selected from Isfahan, Shiraz, and Tehran state universities. A validated questionnaire and group interviews were utilized to collect the data of this study. The data was analyzed and described through descriptive statistical procedures. Based on the results, Freshmen TEFL students considered the UEE mainly unreliable, while they held positive views towards its validity based on the modules of Messick's framework. The overall analysis of the results pointed to the disempowerment role of the test. Furthermore, the obtained results underscored the necessity of more performance-based modes of assessment, such as using portfolios to be included in UEE. The results of the study bear useful implications for high-stake test constructors and policymakers.

Keywords: Assessment, Messick's Framework, University Entrance Exam, Reliability, Validity

1. Introduction

The results of high-stake tests, standardized, can be utilized to make big decisions to affect communities, administrators, teachers, and students (Madaus, 1988). Moreover, they can be used to select and place students in the world (Alderson & Wall, 1993; Choi, 2008).

According to Yildirim (2010), these tests indicated the results as standards to specify if applicants use the programs they need. So, it is not interesting that the results of the studies on high - stake tests indicate a significant effect on the paradigms of teaching and systems of education in different countries (Lombourdi, 2014). Moreover, they make the curriculum narrower (Tsangaris, 2011); teachers employed those changes in their teaching methods (Spratt, 2005; Wall, 2005). Cheng et. al (2004) pointed to the effect of high_ stake tests on making changes in individual learning strategies and course results.

Recently, high stake test effects have drawn the attention of investigators in various instructive contexts. It has also been regarded as one of the central research lines in foreign or second language educational settings (Xie, 2015; Zhang,2014). Large-scale evaluations are the types of outer assessment managed to huge quantities of students for an assortment of reasons (de Lange, 2007). Normally, yet not really, the high-stakes tests are planned to gauge singular accomplishment (e.g., de Lange, 2007; Popham, 2001).

A research study by (Madandar Arani et al., 2012) stated that the assessment system in Iran faces serious problems. Farhadi and Barati (2009), in another study, researched the language assessment policy in Iran. He accepted that the accomplished instructors compose equal tests for each topic standard including the unknown dialect every year. Regarding the strategies of investigating in the exam, it is stated that the only strategy that the ministry of education uses to promote the exam quality is analyzing test items after administration. No written report available was not found regarding these procedures. Khoii (1998) consider both the English subtests' qualitative and quantitative evaluation in his study by using the Rasch model. He proved that there are not satisfactory consequences in reliability and validity of state university entrance examinations. So, there is a need for creating basic fluctuations in the assessment of educational methods. According to Hasani (2005), The emphasis was placed on the change from quantitative assessments to qualitative ones, besides, it was highlighted to replace summative ones with formative evaluations.

The university entrance examinations have been criticized for a long time; in particular, the English tests have received a great deal of criticism for their content from the

investigators' views (e.g., Brown, 1997). Standardized tests involve some problems for several reasons. Norm-referenced tests which have been used to be the much broader curriculum content, cover barely 50% of textbook content, (Freeman et al., 1983). Standardized tests measure just an exceptionally restricted scope of abilities, as opposed to what is educated in the classroom. The questions based on multiple-choice form addresses learning as recollecting confined bits of knowledge, regulations, and strategy. That test-taking guidance aptitudes just raises scores and uncovers the number of components outside of substance information that decide the outcomes.

Further examination on test predisposition unmistakably mirrors the race, class, and sexual orientation advantage managed the cost of certain gatherings in such circumstances. Truth be told, a few instructive antiquarians have connected the underpinnings of government sanctioned testing to the bigoted genetic counseling's development (Sack, 1999). Samuel Messick's all-inclusive record of validity and validation came to rule the instructive and mental estimation and evaluation scene of the 1980s and 1990s. Impelled by Loevinger (1957), created and enunciated by Messick (1989), and embraced through the help of critical partners including (Guion, 1976; Gulliksen, 1950), the quintessence of legitimacy came to be perceived as being essentially a unitary idea. Messick's milestone composition on validity distributed in the reading material Educational Measurement (Messick, 1989) spoke to the finish and articulation of a change in perspective towards a bound together perspective on validity as expressed in the portrayal of current build validity.

Nevertheless, from pupils' viewpoints, assessment is not an empowering experience. The objects of assessment are students, involved in the process of assessment and its decisions (Aitken, 2012; Bound, 2007; Shohamy, 2001, 2007). However, graduation or acceptance to further education as decisions based on these assessments have crucial results on pupils (Bound, 2007; Shohamy, 2001; Virta, 2002).

This study investigated TEFL freshman university students' perspectives on the role of disempowerment of assessment in the university English entrance exam. It also tried to reveal the characteristics of the university entrance exam and the role of assessment in such a high-stake test. It attempted to find out Iranian TEFL students' point of view about the validity and reliability of high-stake tests regardless of many types of research investigating the testing impact on teaching EFL and learning it within various settings. As a research gap,

the empirical study did not focus on the perceptions of the UEE stakeholders about the impact of this high-stakes test on teaching and learning English in Iran.

2. Literature Review

The historical backdrop of the foundation of current instruction with western style focuses on Iran (Persia) dates back to the establishment of Dalarfonun in 1851—founded because of endeavors of the imperial vizier Amir Kabir, that highlighted the innovations of Iranian specialists and their significant role in numerous fields of science. The Islamization of the instructional framework happened after the revolution in 1979.

2.1. University Entrance Exam in Iran

The exploration contemplates done in the previous twenty years or so show that the weaknesses of the University Entrance Examinations in Iran are plentiful. In such manner, the examinations of Yarmohammadi (1986) in which he referenced that the issues of the selection test in state colleges in Iran are colossal can be alluded to. Additionally, Farhadi (1985) broke down the tests of 1983 to 1985 and discovered little correspondence between the way the materials are instructed to the understudies and the way understudies are tried on them .

In instructive composition, there is a general understanding that tests have washback impacts. The wash back impacts insinuate the positive or negative effect that tests have on teaching or learning (Alderson & Wall, 1993; Hughes, 1989). In the past years, tests, particularly high-stakes tests, have been used to progress curricular and enlightening change and get important washback (Cheng, 1997, 2005; Wall, 2000; Saif, 2006). In'nami and Koizumi (2011) led an SEM concentrate on the factor structure of the tuning in and perusing understanding areas of the Test of English for International Communication (TOEIC). They tried a higher-request, a connected, an uncorrelated, and a unitary factor model. The outcomes upheld the related calculate model which in turn underpins the distinguishableness of language capability.

Moreover, the consequences of the multigroup examination proposed the invariance of the related model across various examples. Nonetheless, In nami and Koizumi (2011) examined the Test of English for Academic Purposes (TEAP) and contrasted it and the TOEFL test. Utilizing corroborative factor investigation, they tried four models (unitary,

corresponded, open profitable and higher-request factor model) and found that the higher-request factor model shows the best fit example of this model. The consequences of their investigation demonstrated that there is a cozy connection between TEAP and TOEFL tests and it was proof for the develop legitimacy of this high-stake test.

2.2 .Related Studies

Barati and Ahmadi (2010) examined gender orientation and critical differential thing working (DIF) on the unhitched male's UEE for the candidates into English projects. The investigation used a one-boundary IRT model with an example of about, 36000 test-takers who sat the test in 2004. The discoveries of their investigation affirmed the presence of DIF in a portion of the things of this high-stakes test. Also, utilizing the Rasch model, Ravand and Firoozi (2016) explored the build legitimacy of the 2009 rendition of the Master's UEE for the candidates into English projects. They found that the test in general didn't show uni-dimensionality. Subsequently, they chose to dissect various areas of the test specifically perusing, sentence structure, and jargon independently. As indicated by the creators, the absence of the invariance of the individual measures was another bit of proof against the build validity of the test. However, to the best knowledge of the authors, there have been very few validation studies on the Ph.D. UEE (e.g., Ahmadi et al., 2015; Alibakhshi & Ghandali, 2011). Ahmadi et al. (2015) implied a concurrent triangulation mixed method research study to check the reliability and validity of the Ph.D. UEE in English subset based on Kane's (1992) argument model and Bennett's (2010) action theory. The result of their study indicated that the validity and reliability of this high-stakes test were under the question regarding the test takers' dissatisfaction with test administration conditions including the test venue, testing time, and difficulty level of the IPEET items. Moreover, the results of Logistic Regression (LR) showed 12 items of this high-stake test were flagged for DIF.

In any case, to the best information on the authors, there has been not many approval concentrates on the Ph.D. UEE (Ahmadi et al., 2015; Alibakhshi and Ghandali, 2011). Ahmadi et al. (2015) directed a simultaneous triangulation blended strategy research to check the reliability and validity of the Ph.D. UEE dependent on Kane's (1992) contention paradigm and Bennett's (2010) hypothesis of activity. The aftereffect of their investigation demonstrated that the validity and reliability of this high-stakes test were under the inquiry in regards to the test takers' disappointment with test organization conditions including the

test scene, testing time, and trouble level of the IPEET things. In addition, the consequences of Logistic Regression (LR) demonstrated 12 things of this high-stake test were hailed for DIF.

In another investigation, Gumaa Siddiek (2010) did an examination to investigate the highlights of the Sudan School Certificate English Examination (SSC) from the points of view of substance legitimacy and exhaustiveness. The aftereffects of this investigation demonstrated that SSC English Examination structures are incomprehensive and need content legitimacy. The creator asserted that they are capability tests as opposed to academic normalized accomplishment assessment, and, accordingly, have negative discharge in language instruction improvement in Sudan .

Notwithstanding, contemplates that examine EFL college understudies' discernments on utilizing portfolios appraisal in EFL settings seem, by all accounts, to be restricted. Along these lines, this examination intended to explore EFL Iranian understudies' discernments about the advantages of keeping portfolios, the difficulties that they confronted and their test inclinations. In this manner, this study investigation might be useful to remove the gap in the writing on portfolio appraisal in the EFL context based on the perspectives of the college understudies. Thinking about the targets of this exploration, the specialists have figured the accompanying examination questions:

RQ1. Do TEFL students consider this Exam a valid and reliable way of assessing their English skills?

RQ2. What is the overall experience of Iranian TEFL Freshman undergraduate students about the university English Entrance Exam?

RQ3. What predicts disempowerment in the university English Entrance Exam, and how assessment disempowerment manifests itself?

3. Methodology

3.1. Design and Context of the Study

By using the tenets of the mixed methods paradigm in this study, the instruments were used to collect the data and fulfill the aim of this study. The mixed methods research (MMR) framework concludes the strength and reduces the weaknesses of both quantitative and qualitative data (Johnson & Onwuegbuzie, 2004).

According to Creswell and Plano Clark (2011), the convergent parallel design was chosen among different mixed methods. Regarding the main strength, this design makes new explanations, questions, and hypotheses to emerge (Wolff et al., 1993). Besides, it may confirm and back up the qualitative scales (Creswell & Plano Clark, 2012). The semi-structured interviews and the open_ ended questionnaires, made by research, were used to gather the data. The data, drawn from qualitative and quantitative methods, were combined, compared, and explained.

3.2. Participants

One hundred B.A TEFL freshman students, studying in teaching English as a foreign language, as participants of this study were selected from state Universities of Isfahan, Shiraz, and Tehran. They were first-year students who passed the university English entrance exam successfully. The participants were 67 female and 33 male EFL students who were between 18 and 21 years old. In this study, all the participants were selected conveniently and based on their availability and willingness to participate. (Table 1)

Table 1.

Demographic Background of the Participants

No. of the Students	100
Gender	Male 33
	Female 67
Native Language	Persian
Age Range	18 to 21
Field	English-Literature (20), Translation Studies(20), and TEFL(60)
Universities	Isfahan, Shiraz, and Tehran
Degree	Freshmen

3.3. Instruments

The research instruments of this study included questionnaires and semi-structured interviews.

3.3.1. Questionnaire

This questionnaire was designed and validated by the researcher. It consisted of 40 questions. The questionnaire items covered the following topic areas: the major reasons for the National Organization of Educational Testing which administer the UEE, the main characteristics that have been observed in the English section of exam papers of UEE in recent years, the extent to which disempowerment manifests itself in English section of UEE in Iran, and finally, validity and reliability issues regarding Mesick's framework.

To the extent that the validity of the student's questionnaires is considered, three language experts reviewed the questions so as to determine their appropriateness in terms of the content and language and if they address the objectives of the study adequately. Moreover, a pilot study comprising 30 participants who were asked to fill out the questions, to found the research feasibility. This pilot study helps to predict the potential problems and to examine the reliability of students' questionnaires.

By calculating Cronbach's alpha, internal consistency reliability was estimated for the subscale of the teachers' questionnaires according to the following Cronbach's alpha reliability classification (George & Mallery, 2003):

("_ > .9 – Excellent, _ > .8 – Good, _ > .7 – Acceptable, _ > .6 – Questionable, _ > .5 – Poor, and _ < .5 – Unacceptable").

3.3.2. Interview

Therefore, some general questions were developed to include the main themes of the respective research questions. The participants were qualitatively interviewed to draw their perceptions of high-stakes tests which in this study is, the Iranian Nation-Wide University Entrance Exam's validity and reliability, the appropriateness of this exam to measure all English skills and their overall experience of university English Entrance Exam, the determining factors in the prediction of disempowerment in University English Entrance Exam assessment, and how assessment disempowerment manifests itself.

3.4. Data Collection Procedure

The first stage in doing the research was looking for 100 or more TEFL students who volunteer to participate in the research. The research goals were briefly described to inform participants. They were asked to give the answers to the questionnaire items. The relevant instructions were explained to the participants and problematic items were clarified to

conserve reliability against any unsystematic variance. The questionnaire had two sections. In the first part, the respondents provided their demographic information including their age, gender, and the place where they were studying. The second part, which included items such as, the usefulness of different kinds of assessment, students' goal orientation; empowerment in assessment processes, and the frequency of different kinds of methods. The participants were required to indicate the degree to which they applied the principles in their English language classes, based on their frequency.

The questionnaire consisted of two parts. The first part focused on the demographic information and the second part extracted information about assessment methods and their various types, the goal orientation of pupils, the empowerment of assessment procedure, the frequency of various method types.

Interviewees asked 28 participants several questions to cover the significant themes of queries of the research regarding their overall experience of the university English Entrance Exam, the determining factors in the prediction of disempowerment in the University English Entrance Exam assessment, and how assessment disempowerment manifests itself. However qualitatively, it was suggested to investigate the motivation issues by using one-to-one interviews. Since the semi-structured interviews provide more meaningful opportunities to comment on the research questions.

4.Results

Qualitatively analyzing the data of the Student Perceptions' Assessment Questionnaire (SPAQ), a descriptive research design was used to meet the objectives. Quantitative data were analyzed using the SPSS program.

4.1. Results for the First Research Question

It might be recalled from previous sections that the first research question of the study was: Do EFL students consider the UEE a valid way and also a reliable way of assessing their English skills? In order to find the answer to the question, here, the data obtained from the questionnaire tapping on the students' attitudes about the reliability and validity of the English subtest of the UEE were used and analyzed, as presented in what follows.

4.1.1. Reliability

In an attempt to examine the reliability of the English subtest of the UEE, the relevant questionnaire data are summarized and presented Table 2.

Table 2.

EFL Students' Perceptions of Reliability of the UEE English Subtest

Major Themes	%
Items are difficult, small in font size, and too many in numbers.	9
There is over emphasis on vocabulary, grammar knowledge, and reading comprehension.	16
If you take the test again next time you will not receive the same result.	32
Responses are confusing especially for vocabulary items.	8
Appropriate administration can affect test-takers' performance positively.	12
The test format is not satisfactory and makes students confused and exhausted.	18
There is adequate item structure variety.	5

In Table 2, the factors threatening the reliability of the test are displayed. It could be observed here that 9% of the respondents believed that the items were difficult, small in font size, and too many in numbers. As a case in point, the length is indisputably one of the factors which can affect the reliability of a test. It can also be found that 16% of them contended that the test was unbalanced as there was an emphasis on vocabulary, grammar, and reading comprehension in the test. A striking percentage of 32% believed that the test results and the test setting characteristics are not consistent over various administrations of the test. Other responses which led to the major themes included the confusing nature of responses, the unsatisfactory format of the test, and its being exhausting. Thus, the test could be arguable evaluated as a not reliable one. Tables 2 and 3 show whether the students differed in their attitudes toward the reliability of the test or whether the differences among their attitudes were infinitesimal:

Table 3.

Students' Attitude towards Reliability of the UEE: Frequencies, Percentages, and Std. Residuals

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree	Total
Count	10	56	24	6	4	100
%within Group	13.4%	44.3%	35.6%	4.7%	3.2%	100%
Std. Residual	0.9	-0.2	0.1	-0.5	-0.3	

In Table 3, none of the Std. Residual values were found to be beyond the ranges of ± 1.96 . That is to say, there were not any significant differences between the students' attitudes towards the reliability of the UEE. In addition, to put the results on solid ground, the results of the chi-square in Table 4 were considered.

Table 4.

Students' Attitude towards Reliability of the UEE: Analysis of Chi-Square

Reliability	Test	Value	df	Sig(2-tailed)
Chi-Square	Pearson	1.909	3	.591

The results of the chi-square analysis ($\chi^2(3) = 1.90, P > .05$) in Table 4 showed no significant differences between the students' attitudes towards the reliability of the UEE.

4.1.2. Validity

Like what was done above, first, the major themes in the results regarding the validity of the test are presented in Table 5. Statistical analyses indicated if significant differences were found among the attitudes of students towards the validity of the test or not.

Table 5.

EFL Students' Perceptions of Validity of the UEE English Subtest

Major Themes	%
Inferences drawn from scores of the test are dependent on the empirical evidence sources which have been adequately observed in UEE.	28
The test score inferences are associated with considerations impact of tests that have been adequately considered in UEE.	22
The design decisions extracted from the empirical evidence interpretation reflected in the test relevance or the test usefulness.	32
The decisions of the design were extracted from the interpretation which comes from the empirical evidence to increase and foresee the social justification and political defensibility of utilizing the test.	18

In Table 5, the issues affecting the validity of the test are presented. These were in fact traced in the responses of the participants and put together in the present form in this table. These results, by and large, point to the validity of the test. As a case in point, 32% of the

respondents believed that the design decisions derived from the interpretations of empirical evidence are shown in the domain relevance of the test. In the same vein, other defining characteristics of the test validity attracted responses ranging from 18% to 28% of the participants in the study. Now, Table 6 will show whether the students were all of the same attitudes toward the validity of the test, or whether there were significant differences among them.

Table 6.

Students' Attitude towards the Validity of the UEE: Frequencies, Percentages and Std. Residual

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree	Total
Count	32	46	11	7	4	100
% Within Group	28.0%	40.7%	17.6%	7.3%	5.2%	100%
Sd. Residual	0.4	-0.1	0.1	-0.5	0.8	

None of the Std. Residual values in Table 6 was beyond the ranges of ± 1.96 . That is to say, there are not any significant differences between the students' attitudes towards the validity of the UEE. Even more robust evidence for this conclusion can be found in the results of the chi-square test, which are presented in the following table.

Table 7.

Students' Attitude towards Validity of the UEE: Analysis of Chi-Square

	Test	Value	df	Sig. (2-tailed)
Validity	Pearson Chi-Square	1.526	3	.676

The results of the chi-square analysis ($\chi^2(3) = 1.56, P > .05$) in Table 7 showed that there were not any significant differences among the students' attitudes towards the validity of the UEE.

To put the results for the first research question in a nutshell, contrary to the index of reliability, students believed that the UEE enjoys acceptable validity characteristics. The aims of the second and third research questions in the present study were to explore what predicted disempowerment in assessment and how they manifested themselves in the UEE English subtest. These issues are mentioned in the following sections.

4.2. Results for the Second Research Question

The second research question of the study was intended to find out what the overall experience of Iranian EFL university freshman undergraduate students was about the UEE English subset. The students' perceptions and ideas are summarized and presented in the table that follows:

Table 8.

Overall Experience of Students about UEE English Subtest

Major Themes	%
The test enables students to communicate effectively and efficiently.	16%
The test to prepare students for their future carrier.	29%
The test narrows the gap between English use in classroom and out of the class.	12%
The test widens the gap between top students and weak ones.	26%
The test encourages students to take an active role in learning.	33%
The test motivates students to use integrated skills.	17%
The test is more concerned with the communicative aims.	13%
The test items are practical and close to real life.	15%
The test has an adequate emphasis on communicative activities.	11%
The test has an adequate emphasis on productive language learning skills.	19%

A good number of the surveyed students (33%) believed that the UEE acted as a good cause to encourage students to play an active role in learning. An approximately similar number of students (29%) maintained that the UEE aimed to prepare students for their future careers. They also believed (26%) that it widened the gap between top and weak students. However, very few students held that it served to prepare students for communicative purposes and/or real-life encounters. Table 9 shows whether the students differed in their opinion or whether there were no considerable differences among them.

Table 9.

Students' Experiences of the UEE: Frequencies, Percentages, and Std. Residuals

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree	Total
Count	13	27	15	29	16	100
% Within Group	13.00%	27.00%	15.00%	29.00%	16.00%	100%
Std. Residual	-0.4	0.7	-0.2	0.11	0.9	

In Table 9, none of the Std. Residual values were found to exceed the ranges of ± 1.96 . Consequently, there were not any significant differences among the students' perceptions of the experiences of the UEE English subtest. To statistically confirm this result, the analysis of the chi-square in Table 9 was taken into consideration.

Table 10.

Students' Experiences of the UEE: Analysis of Chi-Square

	Test	Value	df	Sig(2-tailed)
Overall Experiences	Pearson Chi-Square	1.34	3	.482

The results of the chi-square analysis ($\chi^2(3) = 1.34, P > .05$) in Table 10 revealed that there were no significant differences among the students' perceptions of their overall experiences of the UEE.

4.3. Results for the third Research Questions

The third research question sought to find out what predicted assessment disempowerment, and how assessment disempowerment manifested itself in the UEE English subtest. To find answers to this research question, the responses of the students to the relevant section of the questionnaire were reproduced in Table 11.

Table 11.

The Manifestation of Assessment Disempowerment

Statement	Strongly Disagree	Agree	Undecided	Disagree	Strongly Agree	Mean
Students were assessed on what the teacher has not taught them.	4 6.57%	10 9.58%	6 12.60%	27 23.83%	53 46.43%	4.45
How the way the students were assessed was not similar to what they did in class.	10 5.75%	4 7.67%	7 11.50%	37 23.56%	42 44.52%	4.48
The assessment focused on what the students did not understand.	40 44.93%	39 26.57%	5 13.42%	9 5.75%	10 1.91%	4.92
4.27 The assessment focused on what the students did not memorize.	15 30.41%	5 20.27%	4 22.46%	34 10.13%	45 8.49%	4.27
The assessment did not focus on what the students have done in class.	55 3.83%	24 7.12%	3 10.13%	6 28.76%	15 31.78%	4.56
I don't know how a particular assessment task will be marked.	75 46.30%	23 30.95%	0 12.05%	2 3.28%	0 3.01%	5.04
The assessment did not help students to apply what they know to real-life problems.	77 4.93%	13 4.10%	5 17.80%	5 47.39%	0 22.19%	4.91

Table 11 shows the results of students' perceptions of Assessment Empowerment and Disempowerment Manifestation in UEE English subtest. The results revealed that more than 45% of the participants believed that the students were assessed on what the teacher has not taught them and their assessment was not similar to what they did in class. About 40% of them believed that the assessment did not focus on what the students did not memorize and understand. They also did not have enough knowledge about assessment and its implementation in real-life problems (45%). The calculated overall mean score of the items in Table 4.10 ($M = 4.66$) showed that teachers' teaching methods are not enough modified to meet the demands of UEE. In other words, this shows the assessment disempowerment

prevailing among the students who wish to take the UEE. To find out whether this degree of disempowerment was of statistical significance or not, a one-sample t-test was conducted, the results of which are presented in Table 12.

Table 12.

One-sample t-Test Results for Assessment Disempowerment

	Test Value = 3				95% Confidence Interval of the Difference	
	T	Df	Overall Mean Score	Sig. (2-tailed)	Lower	Upper
Assessment Disempowerment	15.032	6	4.66	.000	1.3910	1.9319

The results of the t-test presented in table 11 showed that the students' perceptions regarding assessment disempowerment reached statistical significance ($p < .05$), leading us to the conclusion that assessment disempowerment was significantly manifested in the UEE English subtest.

4.4. The Results of Interview with the TEFL Students

The questions of the interview were developed to investigate the freshmen EFL learners' attitude toward content of UEE consistency with educational curriculum, different kinds of assessment, the extent that the UEE helped them to use their knowledge of real-life obstacles, and finally their perceptions toward University Entrance Exam in general.

The research participants considered the dictation or pronunciation as an unnecessary part, not included in UEE. So, English teachers in high school may disregard the communicative and productive skills in the classroom setting. They may prepare the students to take a standardized test, So UEE is not related to communicative aims. In addition, the freshmen learners claimed that the consequence of UEE can't identify students' strength and that decisions based on UEE are not fair. In summary, because of the significance of UEE, teachers prefer to teach the test in different sections of the UEE tests rather than focusing on other skills. The findings indicated that most of the students tend to select portfolio assessments rather than traditional tests .

Moreover, they regard portfolios as useful and successful assessment. However, some students did not prefer portfolio assessments because they suffered from demotivation, and diffident to act based on portfolio assessments. Since the traditional tests may have a strong influence on the students; they may be unaware of the nature of portfolios. It is time-consuming to become familiar with alternative assessments, and pupils should promote their skills and knowledge.

5. Discussion

The basic role of this investigation was to anticipate disempowerment in University English Entrance Exam evaluation, and how assessment disempowerment shows itself. The consequences of students' impression of Assessment Disempowerment Manifestation in UEE, uncovered that their assessment was not like what they did in class. All the examinations of this investigation brought about similar ends on the disempowerment of UEE. The students dreaded UEE yet even course tests had a lot of weight or pressing factor for their solace. Hence, sort of evaluation (Cassady, 2010; Hembree, 1998; Knekta, 2017) had a reasonable association assessment disempowerment role.

Additionally, results uncovered, not all that out of the blue, students consider this test a more valid however less reliable method of surveying their English aptitudes. As per Messick (1989), the underlying part of developing validity necessitates that the score revealing the arrangement of some random test should coordinate the structure of the test. As the consequences of the current investigation upheld a three-factor model (Reading, Vocabulary, and Grammar), three unique scores ought to be accounted for. Messick's (1989) generalizability part of developing validity requires the test to quantify similar build across various subpopulations. The investigation results proposed that the test assignments of the UEE reasonably gauges similar development across various subpopulations, and the test-takers' presentation is equivalent. In any case, as an outcome, the primary part of the developed legitimacy of the test stayed under inquiry.

The lone reasonable defense for the current bungle can be the impact of UEE. Ongoing exploration upheld the presence of washback impact and makes a differentiation among different types of impacts (Alderson & Wall, 1993; Brown, 1997). The consequences of this investigation supported the presence a negative washback impact. In this way, it very well

may be contended that UEE applies negative washback impacts on the substance, showing strategy, and test improvement in Iran.

In a fairly comparable investigation to the current undertaking, Ghorbani (2008) explored the washback impact of the UEE on language educators' educational program arranging and guidance. The discoveries of his examination indicated that UEE unequivocally influences the "what of instructing" yet not the "how of education" in Iranian EFL instructors. The discoveries of this examination are as per those of Ghorbani. Both "what and how" of instructing are seen to be exceptionally affected by the UEE structure. Practically all language educators, paying little mind to their showing experience, instructive foundation, sex, school type, and school locations, have seen the negative impacts of the UEE.

To completely abuse the intensity of assessment to fortify the nation's English language training framework, it is trusted that in the extremely not so distant future the instructive specialists choose to remember the assessment of oral and aural abilities for the UEE. In the event that this is done, the UEE will pick up face, substance, and construct validity and educators will ideally center past perusing ability alone. All things considered, the UEE doesn't satisfy its hypothetical objective of testing understudies' capacity to utilize the language inventively for open purposes. Utilizing portfolio evaluation is a methodology which can be utilized as a developmental continuous cycle giving criticism to understudies as they progress toward an objective.

The current study significantly focused on the students' assessment experiences to hear their voices. Aitken (2012), Gustafson's and Erickson (2013) studies support the results of this study. The findings showed that students had various experiences and each individual reacted in different ways. Assessments not only should be done in summative ways but also should focus on the consequences of learning.

6. Conclusion

This study can have a few implications for EFL educators' instructions. Numerous Iranian EFL educators are inexperienced with the unfavorable impacts of the UEE instructions, they attempt to change the system to the prerequisites of that exam. Accordingly, they demand to get the mindful impact of UEE and attempt to limit the negative impacts. In the interim, the aftereffects of this examination may be useful to three groups of

people: a) at the micro level, to instructors and students, as the two components of educating and learning measure, b) at the macro level, to the UEE developers and directors, educational plan originators just as policymakers, particularly the individuals who are more worried about offering experimental help for high stake tests wonder.

With the reintroduction of the ideas of formative assessment and summative assessment, in the new central subjects for fundamental training and UEE considers, research on the philosophy and the formative assessment impacts are additionally investigated. Examination and creating different sorts of techniques of formative assessment in the foreign language setting (cf. Dark et al., 2003) could enormously profit the two educators and their students – without it, there is a danger that formative assessment will remain an idea in the public educational plans yet would never change to be genuinely alive in the classroom .

It is trusted that the discoveries from this exploration and further observational investigations later on will arise and reveal important insight to help instructors and testing specialists furnish a more proper appraisal apparatus with which to choose the future professions and lives of an enormous number of Iranian students .

At last, it is crucial to have a further examination that centered on the students' points of view such as high-stake assessment or classroom-based assessment. Examination including students in the development of assessment and then finding out group-work types of assessment advancement, regardless of whether in the classroom, locally or maybe nationally, as has been the situation in Sweden (Erickson and Åberg-Bengtsson, 2012), could likewise empower, even legitimize, the role of students.

References

- Ahmadi, G. (2004). Situation and role of educational improvement assessment in education and process-oriented learning. Paper presented at the first educational assessment symposium, Tehran: Ministry of Education.
- Ahmadi, A., Darabi Bazvand, A., Sahragard, R., & Razmjoo, A. (2015). Investigating the Validity of PhD. Entrance Exam of ELT in Iran in Light of Argument-Based Validity and Theory of Action. *Journal of Teaching Language Skills*, 34(2), 1-37.
- Aitken, N. (2012). Student voice in fair assessment practice. In C. F. Webber & J. L. Lupart (Eds.), *Leading student assessment*. Studies in educational leadership; ISSN 15 (pp. 175-200). Springer Netherlands.
- Alderson, C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Alibakhshi, G., & Ali, H. G. (2011). External validity of TOEFL section of Doctoral Entrance Examination in Iran: A mixed design study. *Theory and Practice in Language Studies*, 1(10), 1304-1310.

- Barati, H., & Ahmadi, A. R. (2010). Gender-based DIF across the Subject Area: A Study of the Iranian National University Entrance Exam. *The Journal of Teaching Language Skills (JTLS)*, 2(3), 1-22.
- Bennett, R. E. (2010). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead UK: Open University Press.
- Boud, D. (2007). Reframing assessment as if learning were important. In D. Boud, & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term* (pp. 14-25). Routledge.
- Brown, J. D., & Hudson, T. (1996). The alternatives in language assessment. *TESOL Quarterly*, 32, 653-675
- Brown J. B. (1997) Textbook Evaluation Form. *The Language Teacher*, 21 (10), 15-21.
- Brown, R., Pressley, M., Van Meter, P. & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction (with low-achieving second-grade readers, *Journal of Educational Psychology*, 88, 18-37.
- Cassady, J. C. (2010). Test anxiety: Comtemporary theories and implications for learning. In J. C. Cassady (Ed.), *Anxiety in schools: The causes, consequences, and solutions for academic anxieties*, (pp. 5-26). Peter Lang.
- Chapman, D. W., & Synder, C. W. (2000). Can high-stakes national testing improve instruction: reexamining conventional wisdom. *International of Educational Development*, 20, 427-474.
- Cheng, L. (1997). How does washback influences teaching? Implications from Hong Kong. *Language and Education*, 11(1), 38-54. www.scholink.org/ojs/index.php/selt
- Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 146-170). Mahwah, Lawrence Erlbaum Associates.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge University Press.
- Cheng, L., Y., Watanabe J., & Curtis, A. (2004). *Washback in language testing: research contexts and methods*. Mahwah, Lawrence Erlbaum.
- Choi, I. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39-62.
- Creswell, J.W. & Plano Clark, V.L. (2011) *Designing and conducting mixed methods research*. 2nd Edition, Sage Publications.
- de Lange, J. (2007). Large-scale assessment and mathematics education. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 1111-1142). National Council of Teachers of Mathematics.
- Derry, J. (2000). Iran. In D. Coulby, R. Cowen, & C. JONES (Eds.). *World yearbook of education 2000: Education in times of transition*. Stylus Publications.
- Erickson, G., & Åberg-Bengtsson, L. (2012). A collaborative approach to national test development. In D. Tsagari & I. Csépes (Eds.), *Collaboration in language testing and assessment* (pp. 93-108). Peter Lang.
- Farhadi, H. (1985). A study of English language test in national university entrance examination. *Roshd Journal of Language Learning*, 4, 10-15.

Research in English Language Pedagogy (2021) 9(1): 235-257

- Farhady, H., Barati, H. (2009). *Language assessment policy in Iran. Annual Review of Applied linguistics*, 29, 131-141.
- Freeman, D.J., Kuhs, T.M., Porter, A.C., Floden, R.E., Schmidt, W.H., & Schwille, J.R. (1983). Do textbooks and tests define a natural curriculum in elementary school mathematics. *Elementary School Journal*, 83(5), 501-513,
- Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment*. Rotledge.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference 11.0 update* (4th ed.). Boston: Allyn & Bacon
- Ghorbani, M. R. (2008). *Washback effect of the university entrance examination on Iranian pre-university English language teachers 'curriculum and instruction* [Doctoral dissertation]. University of Putra Malaysia
- Ghosgol, A. (2005) A new look at educational improvement assessment and its role on educational reforms, Paper presented at first educational assessment symposium, Tehran: Ministry of Education.
- Guion, R. M.(1976).Recruiting, selection, and job placement.In M. D.Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp.777-828). Rand McNally.
- Gulliksen, H. (1950). Intrinsic validity. *American Psychologist*, 5,511-517.
- Gumma Siddiek, A. (2010).Evaluation of the Sudan school certificate English examination. *English Language Teaching Journal*, 3(2), 37-47. [http:// www.ccsenet.org/elt](http://www.ccsenet.org/elt)
- Gustafsson, J. E., & Erickson, G. (2013). To trust or not to trust? Teacher marking versus external marking of national tests. *Educational Assessment, Evaluation and Accountability*, 25(1), 69-87.
- Hasani, M. (2005) *The handbook of implementing descriptive assessment for 1st, 2nd, and 3rd graders*, Available at: <http://baher.blogfa.com/post-12.aspx>.
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, 58(1), 47-77.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge University Press.
- In'nami, Y., & Koizumi, R. (2011). Factor structure of the revised TOEIC test: A multiple-sample analysis. *Language Testing*, 29(1), 131-152
- Johnson, R.B. & Onwuegbuzie, A.J. (2004) Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33, 14-26. <http://dx.doi.org/10.3102/0013189X033007014>
- Jokivuori, P., & Hietala, R. (2007). *Määrällisiä tarinoita: Monimuuttujamenetelmien käyttö ja tulkinta*. Porvoo: WSOY.
- Kane, M. (1992). "An argument-based approach to validity." *Psychological Bulletin* 112:527–535.
- Khoii, R. (1998). *A qualitative and quantitative evaluation of the English subtests of the entrance examinations of Universities using Rasch Model* [doctoral dissertation]. Islamic Azad University: Science and Research Campus, Tehran.
- Knekta, E. (2017). Are all pupils equally motivated to do their best on all tests? Differences in reported test-taking motivation within and between tests with different stakes. *Scandinavian Journal of Educational Research*, 61(1), 95–111.

- Loumbourdi, L. (2014). *The Power and Impact of Standardised Tests: Investigating the Washback of Language Exams in Greece*. Peter Lang GmbH.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635–694.
- Madandar Arani, A. Farahbakhsh, S & Kakia, L. (2012) Primary schools in Japan and Iran: A comparative perspective on educational assessment system, In N. Popov, C. Wolhuter, B. Leutwyler, M. Mihova & J. Ogunleye (Eds). *Comparative education, teacher training, education policy, social inclusion and history of education*. Bureau for educational services.
- Madaus, G. (1988). The influence of testing on the curriculum.” In J. Tanner (Ed.), *The Influence of Testing on the Curriculum*. University of Chicago Press.
- Messick, S. (1989). “Validity.” In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.), Phoenix, AZ: Oryx Press, pp. 13–103.
- Metsämuuronen, J. (2009). *Tutkimuksen tekemisen perusteet ihmistieteissä: Tutkijalaitos [The fundamentals of research in human sciences: The researcher edition]* (4th ed.). Helsinki: International Methelp.
- Popham, W. J. (2001). Teaching to the test? *Educational Leadership*, 58(6), 16-20.
- Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what we can do change it*. MA: Perseus Books.
- Ravand, H., & Firoozi, T. (2016). Investigating Validity of UEE using the Rasch Model. *International Journal of Language Testing*, 6(1), 1- 23.
- Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing*, 23(1), 1-34.
- Shih, C. (2009). How tests change teaching: A model for reference. *English Teaching: Practice and Critique*, 8 (2), 188-206
- Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, 18(4), 373-391.
- Shohamy, E. (2007). Tests as power tools: Looking back, looking forward. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner & C. Doe (Eds.), *Language testing reconsidered* (pp. 141-152). University of Ottawa Press
- Spratt, M. (2005). Washback and the classroom: the implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9, (1), 5–29.
- Tsagari, D. (2011). Washback of a high-stakes English exam on teachers’ perceptions and practices. Selected Papers from the 19th ISTAL, 431- 445.
- Virta, A. (2002). Arviointi oppimisen ja opetuksen punaisena lankana. In E. Lehtinen & T. Hiltunen (Eds.), *Oppiminen ja opettajuus* (pp. 63-86). Turun yliopisto. Kasvatustieteiden tiedekunnan julkaisuja B: 71.
- Wall, D. (2000). The impact of high-stakes testing on teaching and learning. Can this be predicted or controlled? *System*, 28, 499-509.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching*. Cambridge University Press.
- Wolf, T.V., Rode, J.A., Sussman, J., Kellogg, W.A., 2006. Dispelling “design” as the black art of CHI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '06. ACM, New York, NY, USA, pp. 521–530.

Research in English Language Pedagogy (2021) 9(1): 235-257

- Xie, X. (2015). Methods of Test Validation. In E.Shohamy & N.Hornberger (Eds.). *Language testing and assessment. Encyclopedia of language and education* 7 (pp. 177-196). Springer Science + Business Media.
- Yarmohammadi, L. (1986). A review of English language test of national university entrance examination year 1986. *Shiraz University Journal of Social Sciences and Humanities*, 2 (1), 80-88. In Persian.
- Yildirim, O. (2010). Washback effects of a high-stakes university entrance exam: Effects of the English section of the university entrance exam on future English language teachers in Turkey. *The Asian EFL Journal Quarterly*, 12 (2), 92-116.
- Zhang, G. (2014). Achievement gap in China. In J. V. Clark (Ed.), *Closing the achievement gap from an international perspective: Transforming STEM for effective education* (pp. 217–228). Springer.