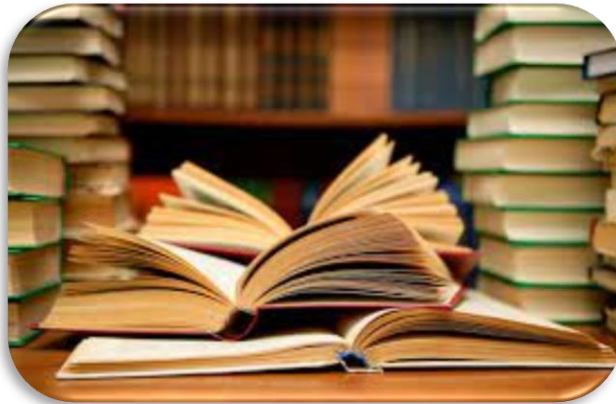


Research Paper



"Fairness Issues in the Iranian EFL Context: Focus on the General English University Entrance Examination (Konkour)"

Sanaz Behboudi¹, Masood Siyyari^{2*}, Gholam-Reza Abbasian³

¹Ph.D. candidate in TEFL, Department of English, Science and Research Branch, Islamic Azad University, Tehran, Iran.

Sanaz.behboudi@srbiau.ac.ir

²Assistant professor of applied linguistics, Department of English, Azad University of Science and Research Branch, Tehran, Iran.

m.siyyari@srbiau.ac.ir

³Member Associate Professor of TEFEL, Department of English Language & Literature, Imam Ali University, Tehran, Iran

gabbasian@gmail.com

Received: 01 January, 2023

Accepted: 10 March, 2023

ABSTRACT

Although many studies have documented the impact of fairness on enhancing students' learning, only a few have been done in English Language Teaching (EFL) in general and in the Iranian EFL context, in particular. Therefore, this study sought to investigate fairness in the context of the Iranian General English University Entrance Examination to see the extent to which it is a fair measure of the candidates' English language ability in terms of admission requirements, format, structure, and content. The researchers developed a questionnaire called the English University Entrance Examination (EUEE), containing two sections: a demographic box and close-ended section with a 5-point Likert-type scale asking respondents to express their opinions. The findings showed that while the majority of respondents agreed that the EUEE met the standards of corporate responsibility and no-test product services, there were significant concerns about its validity, fairness, reliability, test design, equity linking, norming, cut score, test administration, scoring, reporting, and test takers' rights and responsibilities.

Keywords: Fairness; Iranian EFL context; Reliability; Responsibility; Validity

اگرچه بسیاری از مطالعات تأثیر انصاف را بر افزایش یادگیری دانش‌آموزان مستند کرده‌اند، مطالعات کمی در زمینه یادگیری زبان به طور کلی و در زمینه آموزش زبان انگلیسی در ایران به طور خاص انجام شده است. از این رو، این پژوهش به دنبال بررسی انصاف در کنکور در ایران بوده تا میزان منصفانه بودن سنجش توانایی زبان انگلیسی داوطلبان از نظر شرایط پذیرش، قالب، ساختار و محتوا را بررسی کند. محققان پرسشنامه‌ای به نام EUEE تهیه کردند که شامل دو بخش بود: یک بخش مربوط به اطلاعات داوطلبان و بخش دوم آیت‌هایی با مقیاس ۵ درجه‌ای از نوع لیکرت که پاسخ‌دهندگان می‌بایست نظرات خود را بیان کنند. یافته‌ها نشان داد که در حالی که اکثر پاسخ‌دهندگان موافق بودند که کنکور استانداردهای مسئولیت‌پذیری سازمانی را برآورده می‌کند، اما نگرانی‌های قابل توجهی در مورد اعتبار، انصاف، قابلیت اطمینان، طراحی آزمون، معادل‌سازی، پیوند، هنجارسازی، نمره معیار، اجرای آزمون، نمره دهی، گزارش‌دهی و حقوق و مسئولیت‌های شرکت‌کنندگان در آزمون وجود دارد.

کلیدواژه‌ها: انصاف، بافت زبان انگلیسی، روایی، پایایی، مسئولیت‌پذیری

INTRODUCTION

Fairness has received attention from both policy makers and test developers and is considered a key aspect of a test in terms of social justice (McNamara & Roever, 2006). Unfair testing can have negative consequences for examinees and testing institutions that administer the test (Chory, 2007). A fair test is one that is valid for all groups and individuals and makes an equal opportunity for all test takers to demonstrate the skills and knowledge they have acquired (Roever, 2005). As Bachman (1990) suggests, the primary concern in test development and test use is that the interpretations and uses we make of test scores are valid.

This study aims at investigating how much the Iranian university entrance exam is fair. In other words, the main aim of this research study is to investigate if the Konkoor displays different aspects of fairness. Due to the fact that the Konkoor determines examinees' future in terms of their study and career, as well as their personal life, it is imperative that it be free from any kind of bias, and treats all examinees fairly.

Due to the importance of fairness, numerous studies have been carried out and various models have been proposed (e.g., Haertel & Herman, 2005; McNamara & Roever, 2006). However, the studies conducted so far do not yield a compelling account of fairness associated with the Konkoor in Iran. They only propose the general constructs of fairness without going into details of the issue. Thus, the present research aimed to provide a quantified and objectified account of fairness in the Konkoor. This study is important as, in the Iranian context, it is assumed that most language tests in high stakes are not fair, because they do not have validity (Safari, 2016).

A serious pitfall for the Konkoor is that although it has been used as a qualification tool for entering universities in Iran for decades, it has not been seriously and fundamentally revised through these years (Kamyab, 2018). The construct validity of this nationwide exam has been a concern for EFL instructors and education managers (Kamyab, 2008). Thus, findings from the research can guide policy makers and stakeholders in language assessment in detecting identified shortcomings in this regard.

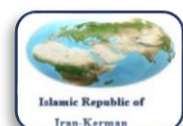
The present research aimed at developing a scale to examine fairness in general English test of Konkoor in Iran. The main question was as follows:

Does the Iranian General English University Entrance Examination (Konkour) meet fairness criteria?

REVIEW OF THE RELATED LITERATURE

The literature has identified and highlighted various aspects of fairness in testing, including, but not limited to, fairness in relation to standardization, test consequences/score use, and item bias (Shohamy & Eldar, 2000). Within the past century or so, the notion of ethics in language testing has been under study by some scholars (Milanovic & Weir, 2004). According to Spolsky (1995), some factors contributed to this notion from 1910s to 1960s, which included "social, economic and political concerns among key language-testing professions in the US and the UK" (As cited in Kunnan, 2018, p. 77). In this regard, Davies (2010) suggested the term 'test virtues' that can be considered as one of the initial suggestions for addressing ethical issues in language testing.

Bias and fairness are closely related but distinct at the same time. Bias is viewed as a statistical feature of the test score or of the prediction based upon those scores. Bias exists when a test involves systematic sources of error in measurement or prediction. The existence of bias can be defined empirically and determined statistically. By examining the data, one can specify the extent to which a test provides bias



measure or bias predictions. On the other side, fairness is associated with a value judgment regarding decisions or actions taken based on the test outcomes. It involves a comparison between the decision that was made and the decision that should have been made.

One way to allay unfairness is multiple assessment through which a lot of related factors can be considered. Another way is to employ multiple phase decision models rather than making irreversible decisions about everyone at the point of testing.

The Test Fairness framework "views fairness in terms of the whole system of a testing practice, not just the test itself" (Kunnan, 2010, p. 45). Therefore, multiple facets of fairness that includes multiple test uses (for intended and unintended purposes), multiple stakeholders in the testing process (examinees, test users, teachers and employers), and multiple steps in the test development process (test design, development, administration and use) are implicated. This model has 5 key features, which are validity, absence of bias, access, administration, and social consequences.

Some researchers have used the differential item functioning (DIF) to detect items whose probability of correct answers differs across different subgroups of a given population, (Chalmers, Counsell, & Flora, 2016). For example, in the EFL context in Iran, Amirian, Alavi, and Fidalgo (2014) investigated whether University of Tehran English Proficiency Test (UTEPT) manifested substantial gender Differential Item Functioning (DIF). They also subjected the flagged DIF to a content analysis to determine underlying sources of DIF. In order to do so, they employed Mantel-Haenszel (MH) and Logistic Regression (LR) as two popular methods of DIF detection. After analyzing the data obtained from 1550 test takers in 2010, they found that "even though 28% of items were initially detected by MH and LR as displaying gender DIF, the effect size of DIF was mostly negligible" (p. 187). In addition to this finding, they conducted a content analysis which indicated that "sometimes it is difficult to hypothesize the linguistic element causing DIF in items" (p.187). In general, they found that humanities-oriented subjects were rated as favoring females and science-oriented subjects were rated as favoring males. Finally, a correlation index of 0.90 manifested that MH and LR produce highly consistent DIF results.

METHOD

Participants and Sampling

The main participants included B.A. Konkoor candidates including 200 students of both sexes (male and female) who had taken Iran's National University Entrance Examination, chosen randomly from university students from different provinces studying at a university in Tehran. This group with the same demographic features as the pilot group filled the questionnaire developed and piloted earlier. Prior to the study, written consent was obtained from all the participants (relevant data are available upon request).

Questionnaire

In order to investigate to what extent the Iranian General English University Entrance Examination (Konkour) is a fair measure of the candidates' English language ability in terms of admission requirements, format, structure, and content, a researcher-made questionnaire was used. The researcher developed a questionnaire based on thorough explorations of research findings and suggestions for



further research found in the relevant literature. The developed questionnaire contained two sections. In the first section there was a demographic box to gather information on the participants' gender, years of language learning experience, and age. The second section, which was the main part of the questionnaire, included close-ended items with a 5-point Likert-type scale asking respondents to read each statement and check the box that most closely represented their opinions, from 1 (strongly agree), 2 (agree), 3 (neutral) 4 (disagree), or 5 (strongly disagree). This questionnaire investigated the extent to which the test is a fair measurement of language ability of the participants.

Data Collection Procedure

First, the group of participants that was used for the pilot phase of the questionnaire was asked to sit for a questionnaire. After piloting, the results of the pilot tests were analyzed statistically to find out if all the items of the test were fine to be used for the actual data collection procedure. All items were checked for their validity and reliability and to see if they are actually testing what they are intended to test. The participants were given as much time they need for completing the questionnaires.

RESULTS

Reliability and Construct Validity of the Fairness Questionnaire

First, we explored the reliability and construct validity of the fairness questionnaire. The fairness questionnaire had 78 items and measured 13 components of Corporate responsibilities (6 items), Widely Applicable Standards (3 items), Non-Test Products and Services (2 items), Validity (8 items), Fairness (7 items), Reliability (6 items), Test Design and Development (9 items), Equating, Linking, Norming, and Cut Scores (9 items), Test Administration (6 items), Scoring (4 items), Reporting Test (7 items), Test (6 items), and Test Takers' Rights and Responsibilities (5 items). The Cronbach's alpha reliability indices of the overall fairness questionnaire and its 13 components showed the overall fairness questionnaire enjoyed a Cronbach's alpha reliability of 0.884. The reliability indices for the 13 components were as follows; Corporate responsibilities ($\alpha = .845$), Widely Applicable Standards ($\alpha = .767$), Non-Test Products and Services ($\alpha = .756$), Validity ($\alpha = .892$), Fairness ($\alpha = .853$), Reliability ($\alpha = .896$), Test Design and Development ($\alpha = .863$), Equating, Linking, Norming, and Cut Scores ($\alpha = .870$), Test Administration ($\alpha = .823$), Scoring ($\alpha = .701$), Reporting Test ($\alpha = .839$), Test ($\alpha = .828$), and Test Takers' Rights and Responsibilities ($\alpha = .827$).

In summary, it can be mentioned that fairness questionnaire and its 13 components enjoyed appropriate reliability indices. That is to say; all reliability indices were higher than the minimum required criterion of .70. The results of EFA indicated that all items loaded under their respective factors; except for items 10 and 11 which had their loadings under the first and eighth factors. All factor loadings for the remaining 76 items enjoyed large effect sizes; i.e., they were higher than 0.50. The results also showed that all 12 extracted factors enjoyed appropriate composite reliability, and convergent validity indices.

The main question raised was whether the Iranian general English University Entrance Examination meets fairness criteria. The results indicated that it did not. In this section, we will address different aspects of fairness in light of the results. Table 1 displays the frequencies and percentages for the first six items measuring "corporate responsibilities".

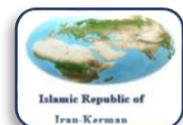


Table 1
Frequencies and Percentages of Corporate Responsibilities

		Fairness					Total
		Strongly disagree	Disagree	Undecided	Agree	Strongly agree	
Helping Quality & Equity	Count	0	25	66	0	167	258
	%	0.0%	9.7%	25.6%	0.0%	64.7%	100.0%
Complying with Laws	Count	0	30	54	0	177	261
	%	0.0%	11.5%	20.7%	0.0%	67.8%	100.0%
Using Funds	Count	21	0	57	95	83	256
	%	8.2%	0.0%	22.3%	37.1%	32.4%	100.0%
Protecting Privacy	Count	29	63	74	65	27	258
	%	11.1%	24.4%	28.7%	25.2%	10.5%	100.0%
Providing Information	Count	33	0	49	80	96	258
	%	12.8%	0.0%	19.0%	31.0%	37.2%	100.0%
Transparency	Count	200	30	0	0	0	230
	%	87.0%	13.0%	0.0%	0.0%	0.0%	100.0%
Total	Count	283	148	300	240	550	1521
	%	18.6%	9.7%	19.7%	15.8%	36.2%	100.0%

The overall results indicated that 52 percent of respondents strongly agreed and agreed with the idea that Iranian general English University Entrance Examination (EUEE) met the standards of “corporate responsibilities”, while 28.3 percent strongly disagreed and disagreed; and another 19.7 percent were undecided. Figure 1 shows the percentages discussed above.

Figure 1
Percentages of Standards of Corporate Responsibilities

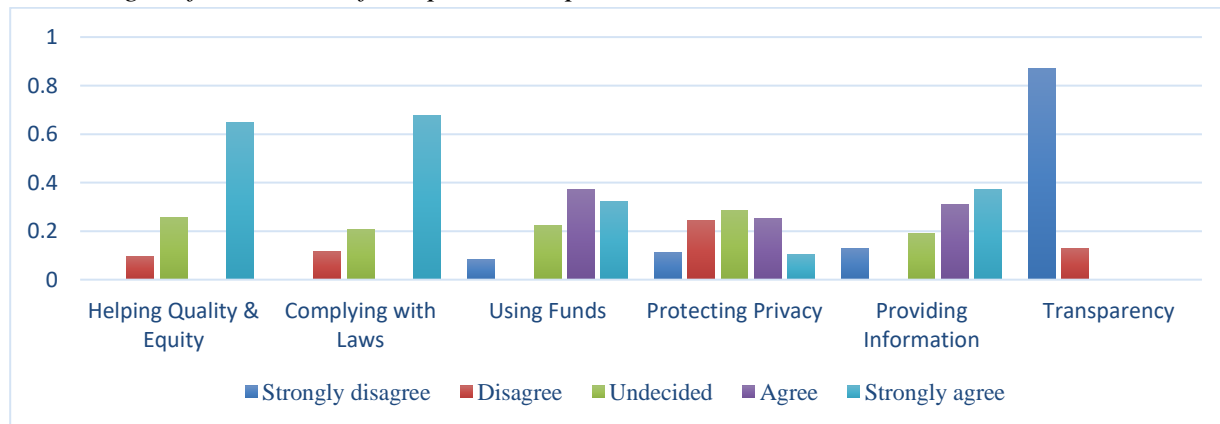


Table 2 displays the frequencies and percentages for the items 7 to 9 measuring “widely applicable standards”.



Table 2
Frequencies and Percentages of Widely Applicable Standards

		Fairness					Total
		Strongly disagree	Disagree	Undecided	Agree	Strongly agree	
Accurate Communication	Count	0	29	57	0	173	259
	%	0.0%	11.2%	22.0%	0.0%	66.8%	100.0%
Decisions are Documented	Count	92	78	82	6	1	259
	%	35.5%	30.1%	31.7%	2.3%	0.4%	100.0%
Qualified Employees	Count	31	0	59	76	91	257
	%	12.0%	0.0%	23.0%	29.6%	35.4%	100.0%
Total	Count	123	107	198	82	265	775
	%	15.9%	13.8%	25.5%	10.6%	34.2%	100.0%

The overall results indicated that 44.8 percent of respondents strongly agreed and agreed with the idea that Iranian general English University Entrance Examination (EUEE) met the standards of “widely applicable standards”. On the other hand; 29.7 percent did not agree with the idea that EUEE meet the applicable standards, and another 25.5 percent were undecided. Figure 2 shows the percentages discussed above.

Figure 3
Percentages of standards of widely applicable standards

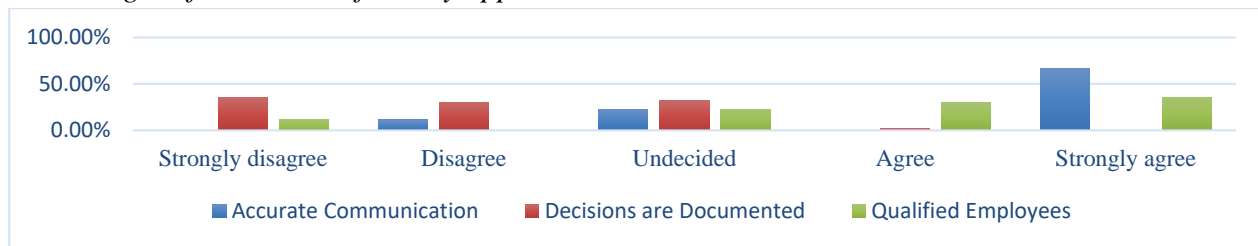
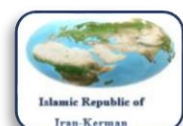


Table 3 shows the frequencies and percentages for items 10 and 11 which measured “no-test product services”.

Table 3
Frequencies and Percentages of No-Test Product and Services

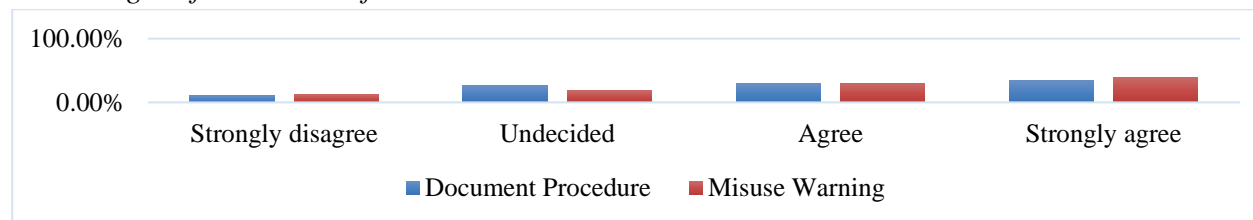
		Fairness				Total
		Strongly disagree	Undecided	Agree	Strongly agree	
Documented Procedures	Count	28	68	76	89	261
	%	10.7%	26.1%	29.1%	34.1%	100.0%
Misuse Warning	Count	33	48	77	101	259
	%	12.8%	18.5%	29.7%	39.0%	100.0%
Total	Count	61	116	153	190	520
	%	11.7%	22.3%	29.5%	36.5%	100.0%



The overall results indicated that 66 percent of respondents strongly agreed and agreed with the idea that the EUEE met the standards of “no-test product services”. On the other hand; 11.7 percent did not agree with the idea that EUEE meet the standards of “no-test products and services”, and another 22.3 percent were undecided. Figure 3 shows the percentages discussed above.

Figure 3

Percentages of Standards of No-Test Products and Services



The fourth component of fairness questionnaire, “validity”, was measured through items 12 to 19. Table 4 shows the frequencies and percentages for the responses given to those eight items.

Table 4

Frequencies and Percentages of Validity

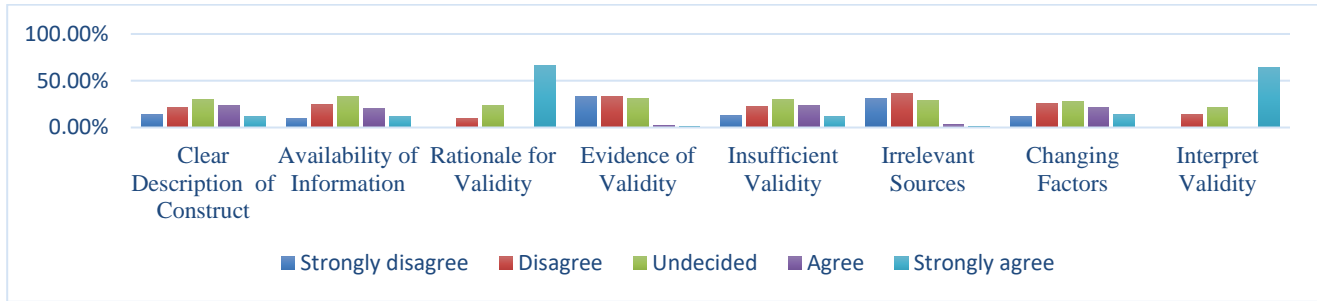
		Fairness					Total
		Strongly disagree	Disagree	Undecided	Agree	Strongly agree	
Clear Description of Construct	Count	36	57	77	60	30	260
	%	13.9%	21.9%	29.6%	23.1%	11.5%	100.0%
Availability of Information	Count	26	63	86	54	31	260
	%	10.0%	24.2%	33.1%	20.8%	11.9%	100.0%
Rationale for Validity	Count	0	25	61	0	170	256
	%	0.0%	9.8%	23.8%	0.0%	66.4%	100.0%
Evidence of Validity	Count	87	87	80	5	2	261
	%	33.3%	33.3%	30.7%	1.9%	0.8%	100.0%
Insufficient Validity	Count	33	59	79	61	30	262
	%	12.5%	22.5%	30.2%	23.3%	11.5%	100.0%
Irrelevant Sources	Count	81	94	75	7	2	259
	%	31.3%	36.3%	29.0%	2.7%	0.7%	100.0%
Changing Factors	Count	31	67	72	56	35	261
	%	11.3%	25.8%	27.7%	21.6%	13.6%	100.0%
Interpret Validity	Count	0	37	56	0	170	263
	%	0.0%	14.1%	21.3%	0.0%	64.6%	100.0%
Total	Count	294	489	586	243	470	2082
	%	14.1%	23.5%	28.1%	11.7%	22.6%	100.0%

The overall results indicated that 37.6 percent of respondents strongly disagreed and disagreed with the idea that the EUEE met the standards of “validity”. On the other hand; 34.3 percent agreed with the



idea that EUEE meet the standards of “validity”, and another 28.1 percent were undecided. Figure 4 shows the percentages discussed above.

Figure 4
Percentages of Standards of Validity



Items 20 to 26 measured “fairness” of EUEE . Based on the results shown in Table 5, it can be concluded that all respondents disagreed with the idea that, “tests are designed, developed, administered, and scored so that they measure the intended construct and minimize the effects of construct-irrelevant characteristics of test takers”. The results also indicated that 37 percent of the respondents agreed and strongly agreed with the idea that, “judgmental and, if feasible, empirical evaluations of fairness of the product or service are obtained and documented for studied groups”, while 31.9 percent held the opposite view; and 31.1 percent were undecided.

Table 5
Frequencies and Percentages of Fairness

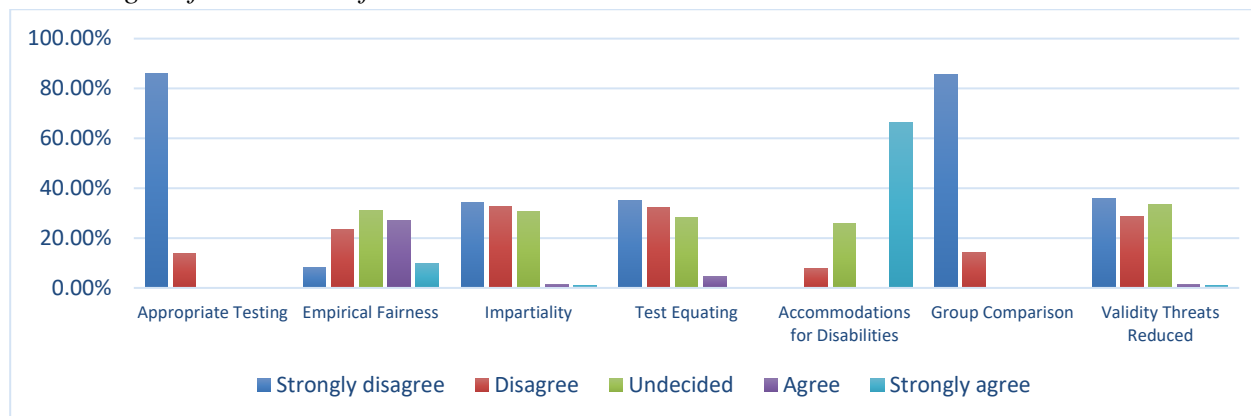
		Fairness					Total
		Strongly disagree	Disagree	Undecided	Agree	Strongly agree	
Appropriate Testing	Count	200	32	0	0	0	232
	%	86.2%	13.8%	0.0%	0.0%	0.0%	100.0%
Empirical Fairness	Count	21	60	79	69	25	254
	%	8.3%	23.6%	31.1%	27.2%	9.8%	100.0%
Impartiality	Count	90	86	80	4	2	262
	%	34.4%	32.8%	30.5%	1.5%	0.8%	100.0%
Test Equating	Count	90	83	73	12	0	258
	%	34.9%	32.1%	28.3%	4.7%	0.0%	100.0%
Accommodations for Disabilities	Count	0	20	66	0	170	256
	%	0.0%	7.8%	25.8%	0.0%	66.4%	100.0%
Group Comparison	Count	194	32	0	0	0	226
	%	85.8%	14.2%	0.0%	0.0%	0.0%	100.0%
Validity Threats Reduced	Count	93	74	87	4	2	260
	%	35.8%	28.5%	33.5%	1.5%	0.7%	100.0%
Total	Count	688	387	385	89	199	1748
	%	39.4%	22.1%	22.0%	5.1%	11.4%	100.0%



The overall results indicated that 61.5 percent of respondents strongly disagreed and disagreed with the idea that the EUEE met the standards of “fairness”. On the other hand; 16.5 percent agreed with the idea that EUEE meet the standards of “fairness”, and another 22 percent were undecided. Figure 5 shows the percentages discussed above.

Figure 5

Percentages of Standards of Fairness



Items 27 to 32 measured “reliability” of EUEE, as shown in Table 6.

Table 6

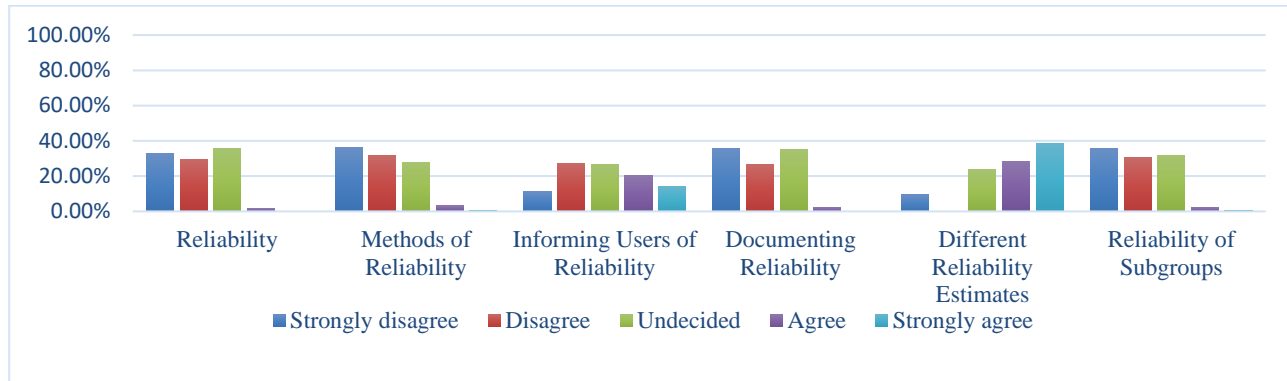
Frequencies and Percentages of Reliability

		Fairness					Total
		Strongly disagree	Disagree	Undecided	Agree	Strongly agree	
Reliability	Count	85	77	92	5	0	259
	%	32.8%	29.6%	35.5%	1.9%	0.0%	100.0%
Methods of Reliability	Count	95	83	73	9	1	261
	%	36.4%	31.8%	28.0%	3.4%	0.4%	100.0%
Informing Users of Reliability	Count	30	71	69	53	37	260
	%	11.5%	27.3%	26.6%	20.4%	14.2%	100.0%
Documenting Reliability	Count	92	68	90	6	0	256
	%	35.9%	26.6%	35.2%	2.3%	0.0%	100.0%
Different Reliability Estimates	Count	24	0	61	72	98	255
	%	9.5%	0.0%	23.9%	28.2%	38.4%	100.0%
Reliability of Subgroups	Count	92	79	82	6	1	260
	%	35.4%	30.4%	31.5%	2.3%	0.4%	100.0%
Total	Count	418	378	467	151	137	1551
	%	27.0%	24.4%	30.1%	9.7%	8.8%	100.0%

The overall results indicated that 51.4 percent of respondents strongly disagreed and disagreed with the idea that the EUEE met the standards of “reliability”. On the other hand; 18.5 percent agreed with the idea that EUEE meet the standards of “reliability”, and another 30.1 percent were undecided. Figure 6 shows the percentages discussed above.



Figure 6
Percentages of Standards of Reliability



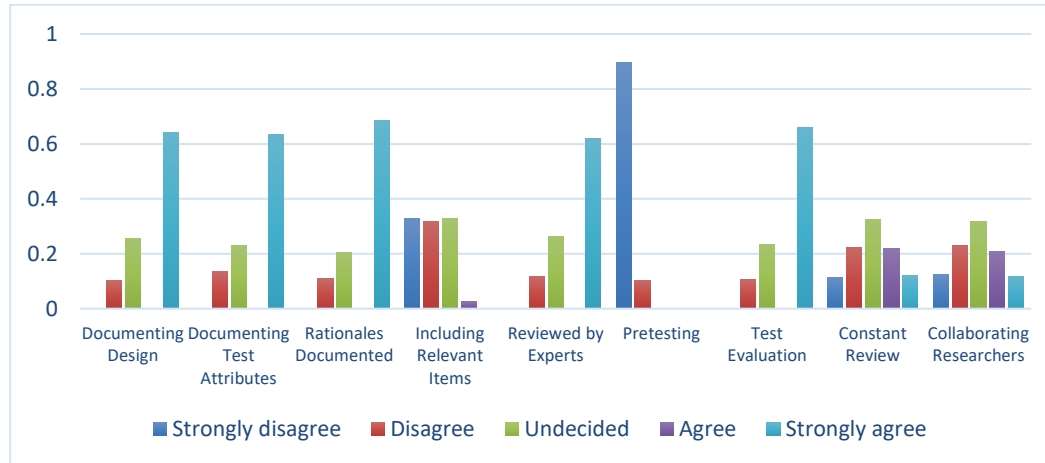
Items 33 to 41 measured “test design and development” of EUEE, as shown in Table 7.

Table 7
Frequencies and Percentages of Test Design and Development

		Fairness					Total
		Strongly disagree	Disagree	Undecided	Agree	Strongly agree	
Documenting Design	Count	0	26	66	0	165	257
	%	0.0%	10.1%	25.7%	0.0%	64.2%	100.0%
Documenting Attributes	Count	0	36	60	0	166	262
	%	0.0%	13.7%	22.9%	0.0%	63.4%	100.0%
Rationales Documented	Count	0	28	53	0	176	257
	%	0.0%	10.9%	20.6%	0.0%	68.5%	100.0%
Including Relevant Items	Count	84	82	84	7	0	257
	%	32.7%	31.9%	32.7%	2.7%	0.0%	100.0%
Reviewed by Experts	Count	0	30	68	0	160	258
	%	0.0%	11.6%	26.4%	0.0%	62.0%	100.0%
Pretesting	Count	207	24	0	0	0	231
	%	89.6%	10.4%	0.0%	0.0%	0.0%	100.0%
Test Evaluation	Count	0	27	60	0	169	256
	%	0.0%	10.5%	23.5%	0.0%	66.0%	100.0%
Constant Review	Count	30	58	84	57	31	260
	%	11.5%	22.3%	32.4%	21.9%	11.9%	100.0%
Collaborating Researchers	Count	33	60	83	55	31	262
	%	12.6%	22.9%	31.7%	21.0%	11.8%	100.0%
Total	Count	354	371	558	119	898	2300
	%	15.4%	16.1%	24.3%	5.2%	39.0%	100.0%

The overall results indicated that 44.2 percent of respondents agreed with the idea that the EUEE met the standards of “test design and development”. On the other hand; 31.5 percent disagreed with the idea that EUEE meet the standards of “test design and development”, and another 24.3 percent were undecided. Figure 7 shows the percentages discussed above.



Figure 7*Percentages of Standards of Test Design and Development*

Items 42 to 50 measured “equating, linking, norming and cut score” of EUEE. The results are shown in Table 8.

Table 8*Frequencies and Percentages of Test Equating, Linking, Norming and Cut Score*

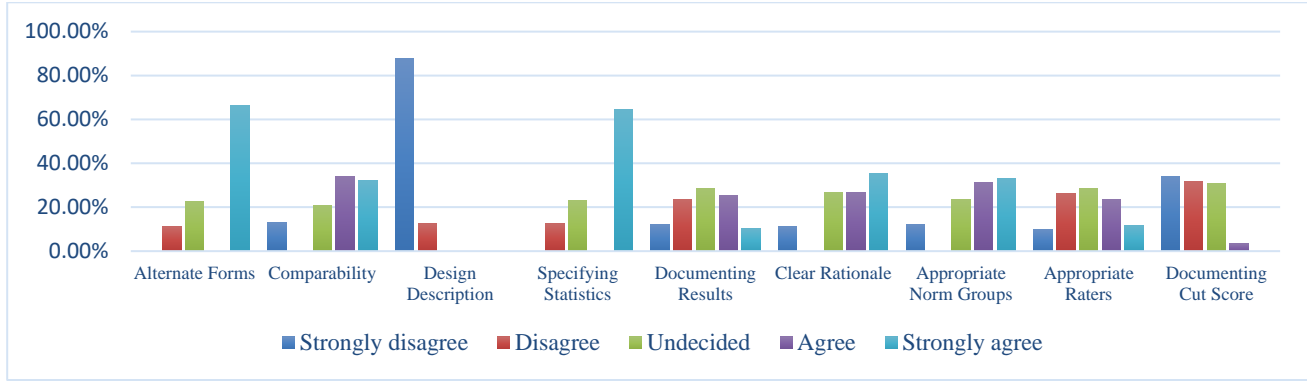
		Fairness					Total
		Strongly disagree	Disagree	Undecided	Agree	Strongly agree	
Alternate Forms	Count	0	29	58	0	172	259
	%	0.0%	11.2%	22.4%	0.0%	66.4%	100.0%
Comparability	Count	34	0	55	89	84	262
	%	13.0%	0.0%	21.0%	34.0%	32.0%	100.0%
Design Description	Count	204	29	0	0	0	233
	%	87.6%	12.4%	0.0%	0.0%	0.0%	100.0%
Specifying Statistics	Count	0	32	59	0	167	258
	%	0.0%	12.4%	22.9%	0.0%	64.7%	100.0%
Documenting Results	Count	32	62	75	67	27	263
	%	12.2%	23.6%	28.4%	25.5%	10.3%	100.0%
Clear Rationale	Count	29	0	70	69	92	260
	%	11.2%	0.0%	26.9%	26.5%	35.4%	100.0%
Appropriate Norm Groups	Count	32	0	62	82	87	263
	%	12.1%	0.0%	23.6%	31.2%	33.1%	100.0%
Appropriate Raters	Count	26	68	74	61	30	259
	%	10.0%	26.2%	28.6%	23.6%	11.6%	100.0%
Documenting Cut Score	Count	87	81	79	9	0	256
	%	34.0%	31.6%	30.9%	3.5%	0.0%	100.0%
Total	Count	444	301	532	377	659	2313
	%	19.2%	13.0%	23.0%	16.3%	28.5%	100.0%



The overall results indicated that 44.8 percent of respondents agreed with the idea that the EUEE met the standards of “equating, linking, norming and cut score”. On the other hand; 32.2 percent disagreed with the idea that EUEE meet the standards of “equating, linking, norming and cut score”, and another 23 percent were undecided. Figure 8 shows the percentages discussed above.

Figure 8

Percentages of Standards of Equating, Linking, Norming and Cut Score



Items 51 to 56 measured “test administration” of EUEE, as shown in Table 9.

Table 9

Frequencies and Percentages of Test Administration

		Fairness					Total
		Strongly disagree	Disagree	Undecided	Agree	Strongly agree	
Administration	Count	209	26	0	0	0	235
Procedure	%	88.9%	11.1%	0.0%	0.0%	0.0%	100.0%
Informing	TestCount	24	66	72	61	33	256
Takers	%	9.4%	25.8%	28.1%	23.8%	12.9%	100.0%
Comfortable	Count	26	0	61	75	95	257
Environment	%	10.1%	0.0%	23.7%	29.2%	37.0%	100.0%
Maintain Security	Count	30	0	67	73	91	261
	%	11.5%	0.0%	25.7%	28.0%	34.8%	100.0%
Eliminating Fraud	Count	29	0	64	77	90	260
	%	11.2%	0.0%	24.6%	29.6%	34.6%	100.0%
Other	DigitalCount	207	28	0	0	0	235
Devices	%	88.1%	11.9%	0.0%	0.0%	0.0%	100.0%
Total	Count	525	120	264	286	309	1504
	%	34.9%	8.0%	17.6%	19.0%	20.5%	100.0%

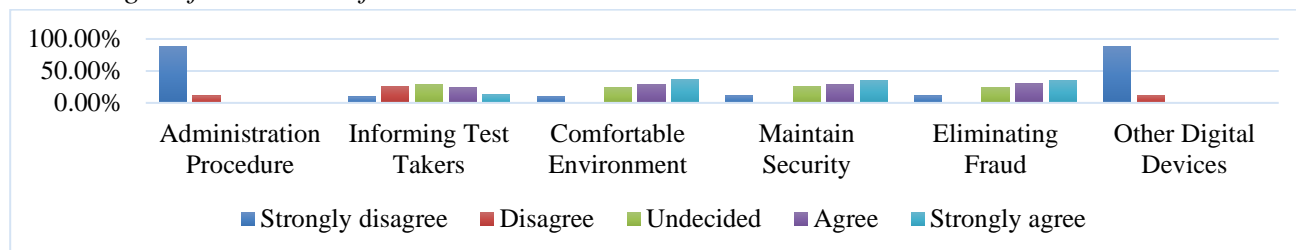
The overall results indicated that 42.9 percent of respondents disagreed with the idea that the EUEE met the standards of “test administration”. On the other hand; 39.5 percent agreed with the idea that



EUEE meet the standards of “test administration”, and another 17.6 percent were undecided. Figure 9 shows the percentages discussed above.

Figure 9

Percentages of Standards of Test Administration



Items 57 to 60 measured “scoring” of EUEE, as summarized in Table 10.

Table 10

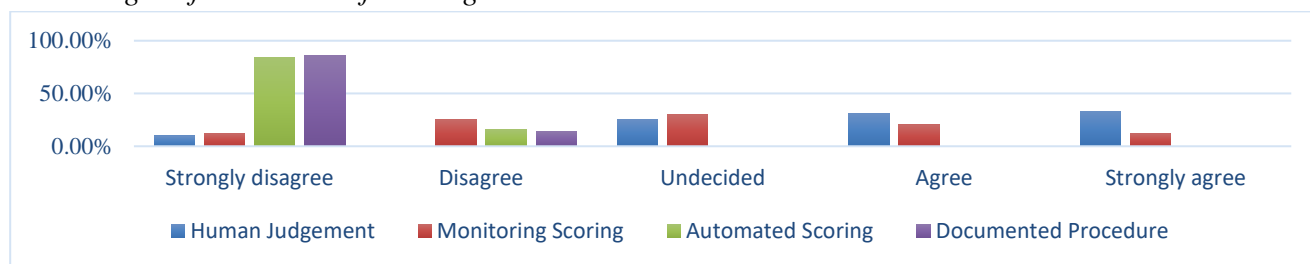
Frequencies and Percentages of Scoring

		Fairness					Total
		Strongly disagree	Disagree	Undecided	Agree	Strongly agree	
Human Judgement	Count	26	0	67	81	87	261
	%	10.0%	0.0%	25.7%	31.0%	33.3%	100.0%
Monitoring Scoring	Count	32	66	78	53	32	261
	%	12.3%	25.3%	29.8%	20.3%	12.3%	100.0%
Automated Scoring	Count	193	37	0	0	0	230
	%	83.9%	16.1%	0.0%	0.0%	0.0%	100.0%
Documented Procedure	Count	198	33	0	0	0	231
	%	85.7%	14.3%	0.0%	0.0%	0.0%	100.0%
Total	Count	449	136	145	134	119	983
	%	45.7%	13.8%	14.8%	13.6%	12.1%	100.0%

The overall results indicated that 59.5 percent of respondents disagreed with the idea that the EUEE met the standards of “scoring”. On the other hand; 25.7 percent agreed with the idea that EUEE meet the standards of “scoring”, and another 14.8 percent were undecided. Figure 10 shows the percentages discussed above.

Figure 10

Percentages of Standards of Scoring



Items 61 to 67 measured “reporting test” of EUEE, as shown in Table 11.

Table 11
Frequencies and Percentages of Reporting Test

		Fairness					Total
		Strongly disagree	Disagree	Undecided	Agree	Strongly agree	
Provide Information	Count	30	64	79	54	34	261
	%	11.5%	24.5%	30.3%	20.7%	13.0%	100.0%
Avoiding Misinterpretation	Count	24	63	84	57	29	257
	%	9.3%	24.5%	32.7%	22.2%	11.3%	100.0%
Misinterpretation of Scale	Count	28	54	81	63	31	257
	%	10.9%	21.0%	31.5%	24.5%	12.1%	100.0%
Appropriate Scale	Count	200	33	0	0	0	233
	%	85.8%	14.2%	0.0%	0.0%	0.0%	100.0%
Stability of Scale	Count	202	30	0	0	0	232
	%	87.1%	12.9%	0.0%	0.0%	0.0%	100.0%
Frame of Reference	Count	89	84	78	8	1	260
	%	34.2%	32.3%	30.0%	3.1%	0.4%	100.0%
Correct Interpretation	Count	0	27	64	0	170	261
	%	0.0%	10.3%	24.5%	0.0%	65.2%	100.0%
Total	Count	573	355	386	182	265	1761
	%	32.5%	20.2%	21.9%	10.4%	15.0%	100.0%

The overall results indicated that 52.7 percent of respondents disagreed with the idea that the EUEE met the standards of “reporting test”. On the other hand; 25.4 percent agreed with the idea that EUEE meet the standards of “reporting test”, and another 21.9 percent were undecided. Figure 11 shows the percentages discussed above.

Figure 11
Percentages of Standards of Reporting Test

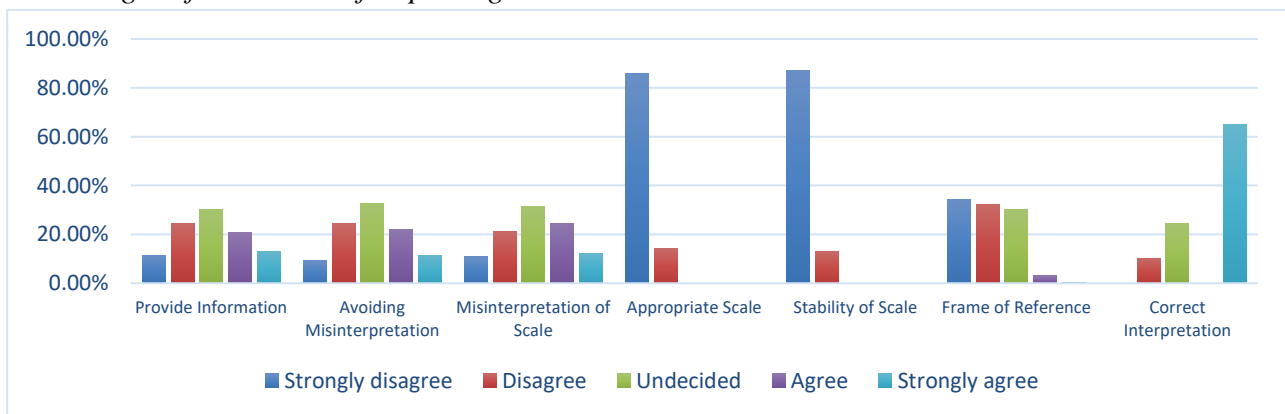


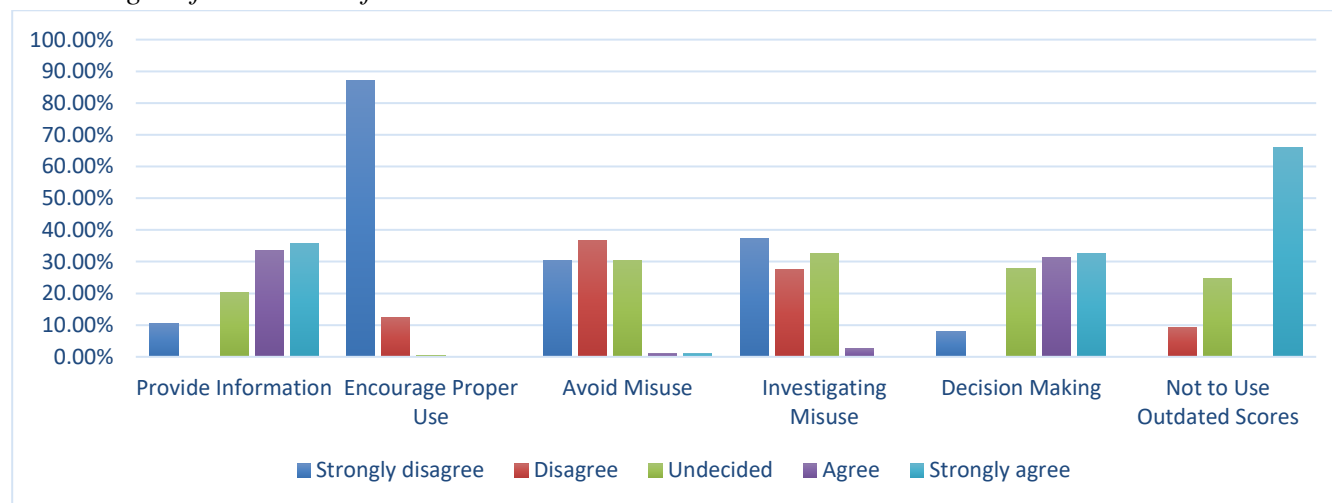
Table 12 shows the frequencies and percentages for “test” criterion which was measured through items 68 to 73.



Table 12*Frequencies and Percentages of Test*

		Fairness					Total
		Strongly disagree	Disagree	Undecided	Agree	Strongly agree	
Provide Information	Count	27	0	52	86	92	257
	%	10.5%	0.0%	20.2%	33.5%	35.8%	100.0%
Encourage Proper Use	Count	203	29	1	0	0	233
	%	87.1%	12.4%	0.4%	0.0%	0.0%	100.0%
Avoid Misuse	Count	79	95	79	3	3	259
	%	30.5%	36.7%	30.6%	1.2%	1.2%	100.0%
Investigating Misuse	Count	97	72	85	7	0	261
	%	37.2%	27.6%	32.6%	2.6%	0.0%	100.0%
Decision Making	Count	21	0	72	81	84	258
	%	8.1%	0.0%	27.9%	31.4%	32.6%	100.0%
Not to Use Outdated Scores	Count	0	24	63	0	168	255
	%	0.0%	9.4%	24.7%	0.0%	65.9%	100.0%
Total	Count	427	220	352	177	347	1523
	%	28.0%	14.4%	23.1%	11.6%	22.8%	100.0%

The overall results indicated that 42.4 percent of respondents disagreed with the idea that the EUEE met the standards of “test”. On the other hand; 34.4 percent agreed with the idea that EUEE meet the standards of “test”, and another 23.1 percent were undecided. Figure 12 shows the percentages discussed above.

Figure 12*Percentages of Standards of Test*

The last criterion of fairness; i.e., “test takers’ rights and responsibilities” was measured through items 74 to 78. The results are shown in Table 13.



Table 13

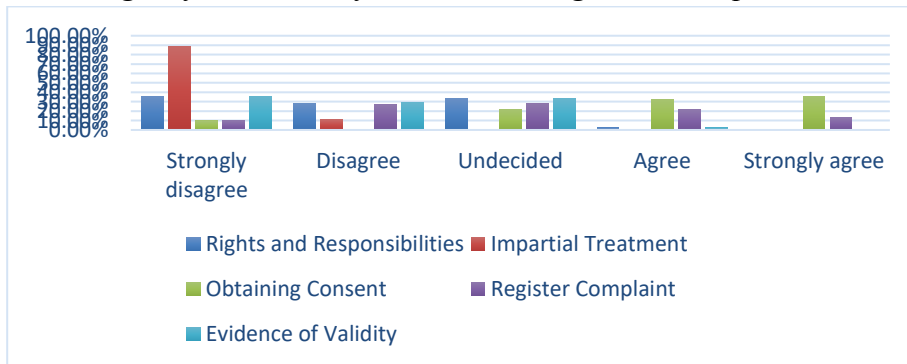
Frequencies and Percentages of Test Takers’ Rights and Responsibilities

		Fairness					Total
		Strongly disagree	Disagree	Undecided	Agree	Strongly agree	
Rights and Responsibilities	Count	92	74	88	6	0	260
	%	35.4%	28.5%	33.8%	2.3%	0.0%	100.0%
Impartial Treatment	Count	204	26	0	0	0	230
	%	88.7%	11.3%	0.0%	0.0%	0.0%	100.0%
Obtaining Consent	Count	25	0	56	84	92	257
	%	9.7%	0.0%	21.8%	32.7%	35.8%	100.0%
Register Complaint	Count	27	70	72	56	35	260
	%	10.4%	26.9%	27.7%	21.5%	13.5%	100.0%
Evidence of Validity	Count	93	77	88	6	0	264
	%	35.2%	29.2%	33.3%	2.3%	0.0%	100.0%
Total	Count	441	247	304	152	127	1271
	%	34.7%	19.4%	23.9%	12.0%	10.0%	100.0%

The overall results indicated that 54.1 percent of respondents disagreed with the idea that the EUEE met the standards of “test takers’ rights and responsibilities”. On the other hand; 12 percent agreed with the idea that EUEE meet the standards of “test takers’ rights and responsibilities”, and another 33.3 percent were undecided. Figure 13 shows the percentages discussed above.

Figure 13

Percentages of Standards of Test Takers’ Rights and Responsibilities



DISCUSSION

This study aimed to evaluate the fairness of the Iranian General English University Entrance Examination (EUEE) by analyzing the responses to a questionnaire. The findings showed that while the majority of respondents agreed that the EUEE met the standards of corporate responsibility and no-test product services, there were significant concerns about its validity, fairness, reliability, test design, equating, linking, norming, cut score, test administration, scoring, reporting, and test takers’ rights and responsibilities.



The study provides important insights into the fairness of the EUEE. These findings suggest that improvements are needed to ensure that the test is reliable, valid, and fair for all examinees, regardless of their gender, school type, or ethnicity. This study has implications for policymakers, test developers, and educators who need to address these issues and ensure that the test meets international standards of fairness.

The findings of the present study suggest that the Iranian General English University Entrance Examination (Konkour) may not meet fairness and reliability standards. These results are consistent with previous studies that have reported concerns about the validity and fairness of high-stakes language exams in different contexts (Alderson & Hamp-Lyons, 1996; Bachman, 1990; Shohamy & Eldar, 2002).

However, it should be noted that the present study was conducted in a specific context and the results may not be directly comparable to other studies. Moreover, the sample size of the present study was relatively small, which may limit the generalizability of the results. Overall, the present study adds to the growing body of research that highlights the importance of evaluating the validity and fairness of high-stakes language exams to ensure that they accurately measure language proficiency and do not unfairly disadvantage certain groups of examinees.

The study discussed above evaluated the fairness and social justice of the Iranian General English University Entrance Examination (EUEE). The study found that there were significant concerns about the validity, fairness, reliability, test design, equating, linking, norming, cut score, test administration, scoring, reporting, and test takers' rights and responsibilities. The study's findings are consistent with previous studies that have reported concerns about the validity and fairness of high-stakes language exams in different contexts. Moreover, the study provides important insights into the validity and fairness of the EUEE, but the results may not be directly comparable to other studies since it was conducted in a specific context and had a relatively small sample size.

The results suggest that the EUEE met the standards of “corporate responsibilities”, “widely applicable standards”, and “no-test product services” according to a majority of respondents. However, the exam did not meet the standards of “validity”, “fairness”, “reliability”, “test administration”, “scoring”, “reporting test”, “test”, and “test takers' rights and responsibilities” based on the responses of more than half of the participants.

These findings are important because they raise concerns about the overall quality of the EUEE and the extent to which it can accurately measure students' English language proficiency. The low scores on the fairness and validity standards are particularly concerning because these are crucial components of any high-stakes exam, especially in the context of university entrance examinations. These results suggest that the EUEE may not be providing a fair and valid assessment of students' language abilities, which could have significant implications for students' educational and professional opportunities.

Overall, the results of this study raise important questions about the fairness, validity, and social justice criteria of the Iranian General English University Entrance Examination. These findings highlight the need for further research and evaluation of the EUEE, as well as potential reforms to ensure that the exam is providing a fair and accurate assessment of students' language abilities.

Fairness is a crucial aspect of any high-stakes examination, especially in the context of university entrance exams like the Iranian General English University Entrance Examination (EUEE). Fairness



ensures that all examinees have an equal opportunity to demonstrate their knowledge and skills, regardless of their background, gender, ethnicity, or school type. In the context of the EUEE, fairness involves examining the extent to which the test accurately measures language proficiency and does not disadvantage certain groups of examinees.

As discussed in the previous section, the study evaluating the EUEE raised significant concerns about the fairness of the exam. According to the responses from more than half of the participants, the EUEE did not meet the standards of fairness. This suggests that some aspects of the test might be biased or disadvantage certain groups of examinees, leading to potential inequities in their results.

Fairness is particularly crucial for underrepresented groups, including individuals from low-income backgrounds, ethnic minorities, and students attending public schools. If the EUEE contains biases or advantages certain groups, it could perpetuate existing social inequalities and limit the opportunities for these students to access higher education.

To ensure fairness in the EUEE, it is essential for policymakers, test developers, and educators to carefully examine the test items, scoring methods, and administration processes. They should identify potential biases and take appropriate measures to address them. This may involve revising certain items, implementing standardized procedures for test administration, and conducting regular fairness evaluations.

Social justice goes beyond fairness and emphasizes the need for equitable opportunities and outcomes for all individuals, regardless of their background. In the context of the EUEE, promoting social justice means creating an inclusive and supportive testing environment that considers the unique circumstances and needs of each examinee.

One of the key aspects of social justice in the EUEE is recognizing the diverse backgrounds of test takers. This involves acknowledging that students may come from various socioeconomic, cultural, and educational backgrounds, which can influence their test performance. By considering these factors, the exam can provide a more holistic and accurate representation of students' language proficiency.

To promote social justice, the EUEE should incorporate mechanisms to accommodate the individual circumstances of test takers. This may include providing reasonable accommodations for students with disabilities or special needs and considering extenuating circumstances that may have affected their preparation or test performance.

Developers of the EUEE should adhere to ethical guidelines and principles throughout the test development process. Transparency in test design, item selection, and scoring criteria is essential for building trust among test takers and the broader community. Test developers should aim to create a test that is not only valid and reliable but also aligns with the principles of social justice.

Regularly evaluating the social impact of the EUEE is essential to identify potential issues related to social justice. This evaluation should include gathering feedback from test takers, educators, and other stakeholders to understand their perspectives on the test's fairness and social justice criteria. Based on the findings, appropriate adjustments can be made to enhance the exam's social impact. Therefore, addressing fairness and social justice concerns in the Iranian General English University Entrance Examination (EUEE) is of utmost importance. By continuously evaluating and improving the test's validity, fairness, and social justice criteria, policymakers and educators can ensure that the EUEE



provides a fair and equitable opportunity for all students to demonstrate their English language proficiency and access higher education.

CONCLUSION

The main research question was aimed at finding if Iranian General English University Entrance Examination (Konkour) meets fairness criteria. In order to test this hypothesis, the questionnaire was run for content analysis. The results suggest that the EUEE met the standards of “corporate responsibilities”, “widely applicable standards”, and “no-test product services” according to a majority of respondents. However, the exam did not meet the standards of “validity”, “fairness”, “reliability”, “test administration”, “scoring”, “reporting test”, “test”, and “test takers’ rights and responsibilities” based on the responses of more than half of the participants. The results of this study can have some implications for teachers, test developers and the mainstream education, especially the Ministry of Education of Iran. One of the implications that can be made from the results of this study is for language teachers. By studying the results of this study, language teachers can become aware of the factors that have impacts on their students' performance in the Konkour, which may eventually lead to their future, especially finding a suitable job. Becoming aware of the shortcomings of the test and the factors that lead to some bias can be a very important factor for improving it by adjusting the expectations towards the test.

REFERENCES

- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language testing*, 13(3), 280-297. <https://doi.org/10.1177/026553229601300304>
- Amirian, S. M. R., Alavi, S. M., & Fidalgo, A. M. (2014). Detecting gender DIF with an English proficiency test in EFL context. *Iranian Journal of Language Testing*, 4(2), 187-203.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Chalmers, R. P., Counsell, A., & Flora, D. B. (2016). It might not make a big DIF: Improved differential test functioning statistics that account for sampling variability. *Educational and Psychological Measurement*, 76(1), 114-140. <https://doi.org/10.1177/0013164415584576>
- Chory, R. M. (2007). Enhancing student perceptions of fairness: The relationship between instructor credibility and classroom justice. *Communication Education*, 56, 89-105. <https://doi.org/10.1080/03634520600994300>
- Davies, A. (2010). Test fairness: A response. *Language Testing*, 27(2), 171-176. <https://doi.org/10.1177/0265532209349466>
- Haertel, E., & Herman, J. (2005). *Historical perspective on validity arguments for accountability testing, CSE Report 654*. Los Angeles, CA: University of California, Los Angeles.
- Kamyab, S. (2008). The university entrance exam crisis in Iran. *International Higher Education*, (51). <https://doi.org/10.6017/ihe.2008.51.8010>
- Karami, H. (2011). Detecting gender bias in a language proficiency test. *International Journal of Language Studies*, 5(2), 27-38.



- Kunnan, A. J. (2010). Test fairness and Toulmin's argument structure. *Language Testing*, 27(2), 183-189. <https://doi.org/10.1177/0265532209349468>
- Kunnan, A. J. (2018). *Evaluating language assessments*. New York: Routledge. <https://doi.org/10.4324/9780203803554>
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell Publishing. <https://doi.org/10.1111/j.1473-4192.2006.00117.x>
- Milanovic, M., & Weir, C. J. (Eds.). (2004). *European Language Testing in a Global Context: Proceedings of the ALTE Barcelona Conference July 2001* (Vol. 18). Cambridge University Press.
- Roever, C. (2005). "That's not fair!" *Fairness, bias, and differential item functioning in language testing*. Retrieved November 18, 2006, from the University of Hawai'i System Web site: <http://www2.hawaii.edu/~roever/brownbag.pdf>
- Safari, P. (2016). Reconsideration of language assessment is a MUST for democratic testing in the educational system of Iran. *Interchange*, 47(3), 267-296. <https://doi.org/10.1007/s10780-016-9276-8>
- Shohamy, E., & Eldar, S. (2002). *High stakes exams and washback: The case of the Bagrut in Israel*. *Assessment in Education*, 9(3), 307-333.
- Spolsky, B. (1995). Measured words: The development of objective language testing.

Biodata

Sanaz Bahboudi is a PhD candidate in TEFL at Science and Research Branch of Islamic Azad University (SRBIAU) in Tehran, Iran. She has taught different general English courses such as writing, reading, and speaking. She has translated more than 10 books. She is keenly interested in conducting research on fairness and social justice and general issues concerning second language acquisition.

Masood Siyyari is an assistant professor of TEFL and applied linguistics at Science and Research Branch of Islamic Azad University (SRBIAU) in Tehran, Iran. He received his Ph.D. in applied linguistics from Allameh Tabataba'i University in Tehran, Iran. Currently, he supervises MA theses and Ph.D. dissertations and teaches MA and Ph.D. courses in applied linguistics and translation studies at SRBIAU. His main areas of research include language assessment and second language acquisition.

Gholam-Reza Abbasian, is a member of the Teaching English Language & Literature Society of Iran (TELLSI) Board of Directors, a presenter at international conferences, an author and a translator of about 15 books. He is an expert in language Testing and Assessment, Research Methods, SLA, and Psycholinguistics, and has supervised about 100 theses and dissertations. He is the internal manager of JOMM, and reviewer of Sage, FLA & GJER and other journals.

