Research in English Language Pedagogy (2024) 12(4): 740-765

DOI: 10.30486/RELP.2024.1120951

Original Research

Detecting Academic Field-based Differential Item Functioning and Differential Distractor Functioning in the Islamic Azad University EPT Employing the Rasch Model

Sarallah Jafaripour[1], *Omid Tabatabaei*[1*], *Hadi Salehi*[1], *Hossein Vahid Dastjerdi*[1]

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

Submission date: 27-05-2024 Acceptance date: 12-08-2024

Abstract

This study aimed to explore Differential Item Functioning (DIF) and Differential Distractor Functioning (DDF) based upon academic fields in the Islamic Azad University English Proficiency Test (IAUEPT). Applying the Rasch model, DIF and DDF were subjected to analysis among PhD candidates belonging to different fields. The 1069 participants were broken down into Social and Human Disciplines (SHD) and Non-Social and Human Disciplines (N-SHD) groups. Findings suggested minimal academic field-related DIF, with only two out of 100 items displaying such effects. Likewise, academic field-specific DDF analysis highlighted disparities in the number of items: one in Choice A, three in Choice B, four in Choice C, and three in Choice D. The results have important implications for individuals involved in the development of high-stakes proficiency tests. Through identifying potential biases in such tests, it becomes possible to promote fairness and equality for all testees. Additionally, the identification of academic field-based DIF can provide insights into potential gaps in the curriculum, revealing areas where learners from different fields of study may face disadvantages.

Keywords: Differential Distractor Functioning (DDF), Differential Item Functioning (DIF), English Proficiency Test (EPT), Rasch Model, Test Bias.

[*] Corresponding Author's E-mail: omidtabatabaei1349@gmail.com

1. Introduction

The concept of assessment fairness which emphasizes equitable treatment and impartial outcomes is widely recognized as a crucial component of assessment validity that is increasingly gaining attention and cannot be overlooked. As stated by Messick (1989), validity involves comprehensively analyzing how well experimental evidence and conceptual grounds uphold the sufficiency and applicability of commentaries and decisions made based off of examination results. DIF is a prominent aspect at assessing the validity of instructive and psychological assessments, happening when two groups with similar abilities show differing levels of proficiency in answering a specific question (Karami, 2012). DIF analysis arose out as a vital constituent in the process of assessment validation owing to its capacity to identify items that may pose a threat to the assessment validity (Ahmadi & Jalili, 2014; Ajideh et al., 2022; Alavi et al., 2012; Ayoobiyan & Ahmadi, 2023; Jafaripour et al., 2024; Zumbo, 2007).

Differential Distractor Functioning (DDF) method, utilized to analyze the responses of examinees, offers insights into group disparities in the selection rates of distractors (Penfield, 2008). This method elucidates the factors contributing to DIF and presents evidence that the origins of DIF can be traced back to the characteristics of correct responses (Adibatmaz & Yildiz, 2020).

DIF is an essential factor to consider when assessing bias, but it alone is not enough to determine bias. In essence, bias is present only when there is an unjustifiable discrepancy in the performance of different groups on a particular item (Ahmadi & Jalili, 2014; Ajideh et al., 2022; Zumbo, 1999). Additional investigations are imperative in order to comprehensively realize the underlying causes of DIF and ascertain the presence of bias. In relation to DDF, several studies have recognized the inherent limitations of employing DDF as the sole method for identifying item bias (e.g., Penfield, 2011; Suh & Talley, 2015; Tsaousis et al., 2018).

Item Response Theory (IRT) has emerged as a prominent model for the detection of DIF and DDF, and its application has gained significant traction in recent years (Moradi et al., 2022; Penton et al., 2022; Sadeghi & Abolfazli Khonbi, 2017). As declared by Ayoobiyan and Ahmadi (2023), the Rasch model is a one-parameter model in IRT in which merely item difficulty is described as a parameter. This model has been identified as highly valuable for assessing the performance of individual test items and validating different types of scales (Sazegar et al., 2021; Stemler & Naples, 2021). As stated by Törmäkangas (2011), the Rasch measurement model offers a notable advantage in that the item characteristic curve is determined by the logistic distribution, which maintains a consistent form across all items. The only variation lies in the location or difficulty on the item calibration scale, as well as the observed probabilities, which differ from item to item.

2. Literature Review

The analysis of DIF holds significant importance in the assessment of tests for the purposes of fairness and equity. It is widely acknowledged as a standardized approach to validate assessments within the broader field of assessment (Penfield & Lam, 2000). According to Zumbo (1999), DIF occurs when there is a statistically significant difference in the likelihood that test-takers from different groups, such as males and females, who have the same underlying ability on the measured construct, demonstrate varying probabilities of accurately responding to a specific item. The ability levels of test-takers are determined by their overall scores on the test. It is important for the DIF analysis of a particular item to be as independent as possible from the DIF analyses of other test items. DIF often indicates the presence of a systematic but irrelevant variance that is being influenced by the test. Additionally, this irrelevant variance is associated with group membership, suggesting that the presence and extent of multidimensionality is influenced by the groups to which individuals belong (Zumbo, 1999).

DIF in tests, whether uniform or non-uniform, has significant implications for ensuring test fairness. Uniform DIF occurs when an item functions differently across all levels of the ability trait, while non-uniform DIF occurs when an item functions differently for specific levels of the ability trait. The detection of DIF is crucial for achieving equitable measurement, as it ensures that all test takers, regardless of their group membership, have an equal opportunity to answer the questions correctly. The identification of DIF items plays a vital role in enhancing the fairness of the test by either modifying or eliminating biased items (Camilli & Penfield, 1997; Clauser & Mazor, 2005).

DDF is another potentially effective approach for detecting DIF. As mentioned by Mapuranga et al. (2008), this approach involves examining the choices made by test-takers who answer an item incorrectly. Through analyzing the differences in response choices, DDF can assist assessment developers in comprehending group disparities in testing.

Furthermore, it can offer a valuable tool for facilitating a substantive and qualitative interpretation of DIF analyses.

DIF can arise from DDF, wherein individuals with the same skill level in one group may have a tendency to choose a particular distractor more frequently than their counterparts in another group. To mitigate item bias, it is important to analyze not only the accuracy of responses but also the behavior of distractors (Kato et al., 2009). It could also work vice versa: an item that displays DIF is expected to show DDF for one or more response options since response rates are dependent on each other (Kato et al., 2009). However, it is conceivable for items to exhibit DDF without displaying DIF. In such instances, the distractors of an item may exhibit a preference towards a specific subgroup within the population, but this preference does not impact the subgroup's ability to respond accurately. As a result, items that exhibit differential functioning across different subgroups may be overlooked if the analysis is solely focused on items demonstrating DIF.

Today, various scholars believe that both DIF and DDF are necessary to validate assessment items. Over the past few years, a multitude of studies were conducted as a means to evaluating the effects of distractors (e.g., Gierl et al., 2017; Hoshino, 2013; Terao & Ishii, 2020). However, a number of studies examining DIF and DDF, some of which are discussed in this review, have produced inconsistent results.

2.1. Previous Research on DIF

In the recent times, there has been a growing scholarly interest in the evaluation of DIF with a particular focus on the examinees' academic fields. Amidst numerous studies on the academic field DIF, Barati et al. (2006) attempted to detect DIF on the English subtest of a standardized assessment. The influence of the testees' academic fields was considered in the study. They found that there were thirty three items on the English subtest that displayed DIF across various sections. Specifically, the vocabulary and word order sections favored students with a background in sciences or humanities, while the grammar, language function, cloze test, and reading comprehension sections had approximately an equal number of DIF items that favored math students and students from sciences and humanities. This outcome is not parallel to an investigation conducted by Alavi et al. (2012) who endeavored to discover DIF in a high-stakes proficiency assessment. In accordance with the assessment of effect size for logistic regression, no substantial effect sizes were detected. The conclusion was that the assessment does not exhibit academic field DIF, and is unbiased towards both the humanities and science and engineering groups.

In their study, Swamanithan and Rogers (1990) conducted an investigation utilizing logistic regression techniques to identify DIF. The researchers differentiated between uniform and nonuniform DIF based on the model's parameters. Additionally, they developed a statistical test to assess whether there is any DIF. The results of investigations related to the simulation specified that the procedure associated with logistic regression exhibited greater amount of effectiveness in detecting nonuniform DIF. Likewise, Kim and Cohen (1995) aimed to liken three processes for identifying DIF. Hence, they used IRT. Besides, they employed Lord's chi-square, Raju's area measures, and the likelihood ratio test. They attempted to examine the correspondence among the three processes via analyzing statistics from two forms of a placement test. The research outcomes publicized great amount of agreement among these techniques.

Takala and Kaftandjieva (2000) conducted a study utilizing IRT to examine gender-related DIF in a vocabulary test designed in L2. The impact of gender on test performance was the goal of the study, as measured by different item composites. The findings revealed that some items had DIF in favor of males and females. Similarly, Alavi and Bordbar (2018) ran a study to examine if there was DIF based on sex in a national university entrance exam. Based on the outcomes, DIF existed between both genders, and it was discovered that the scores have been affected by a group of irrelevant factors.

2.2. Previous Research on DDF

Abedi et al. (2008) attempted to investigate the existence of DDF in a set of reading assessments that are administered to those learners who have special disabilities. The results unveiled the existence of DDF in assessments related to the ninth grade. Furthermore, the findings revealed that items displaying DDF were more frequently observed in the latter half of the assessments rather than the initial half. Their findings are in harmony with the study conducted by Jafaripour et al. (2024) who sought to evaluate DIF and DDF with respect to genders and academic fields utilizing the Rasch model. The results of the DDF analysis on the basis of gender revealed one item for Choices A, B, and C each, and four items for Choice D.

In a study conducted by Penfield (2008), the nominal response model was scrutinized in order to assess if it involves DDF or not. For this aim, odds ratio has been utilized. Via simulation revision, it was uncovered that in different situations, DDF presented popular statistical features. Similarly, Koon and Kamata (2013) conducted a practical investigation into various techniques for detecting DDF. The study utilized the odds-ratio method and uniform distractor analysis to examine the effects of DDF. Data from a statewide mathematics assessment were analyzed to address twofold exploration questions: 1) whether DDF size is the same among all approaches, and 2) whether the impacts of of DDF accommodates to DIF. The results indicated that all three methods offer viable options for enhancing items used in evaluation plans.

While conducting this research, there was limited existing research on DDF, but numerous studies have been carried out to examine DIF. Given the scarcity of studies addressing both DIF and DDF simultaneously, it can be concluded that the current research is quite innovative. As reported by Green, Crone, and Folk (1989), if various groups have a preference for different incorrect responses to an item, which are commonly referred to as distractors, then it suggests that the item holds different interpretations for each group. Items that have different meanings for different groups can be considered biased in a fundamental way. When an assessment exhibits significant DDF, it provides indications that the assessment items possess varying meanings for different groups. Accordingly, this suggests that the interpretation of assessment scores cannot be uniform across different groups. This study stands out because it examines both DIF and DDF to assess the test scores of PhD candidates across different academic fields who take the IAUEPT. This is why the findings from the current research can offer valuable perceptions of the validity of the IAUEPT.

Numerous studies have explored DIF (e.g., Alavi et al., 2012; Barati et al., 2006; Bowles, 2022; Khalaf & Omara, 2022), but this research uniquely targets both DIF and DDF. It is evident that solely analyzing DIF is not sufficient to validate a high-stakes proficiency assessment (Penfield, 2010). To address this gap, the Rasch model was applied in this research to assess both DIF and DDF within the IAUEPT. Consequently, this study is designed to answer the following research questions focusing on academic field-specific DIF and DDF within the IAUEPT:

Q1. Does examinees' academic field lead to significant DIF in the IAUEPT?

Q2. Does examinees' academic field lead to significant DDF in the IAUEPT?

3. Method

3.1. Design of the Study

This investigation was quantitative comparative which involved an experimental research design for responding to the research questions and investigating DIF and DDF on the IAUEPT based on the academic fields of the examinees. To this end, a set of statistical analyses were run to estimate the reliability of the IAUEPT and to analyze the test items for detecting academic field-based DIF and DDF respectively.

3.2. Participants

The current study involved a total of 1069 PhD candidates from a set of academic fields. The spring 2021-2022 edition of the IAUEPT was examined to identify DIF and DDF across diverse academic fields. The purpose of the test is to determine the most suitable exit program for PhD candidates. The age range of the participants was between 23 and 49 years. Table 1 displays a breakdown of different fields of study included in the test.

Table 1.

Details about the Academic Field within the Population

Group	Total	Female	Male
1. Social and Human Disciplines	593	393	200
Humanities	176	132	44
Psychology	207	148	59
Educational Sciences	210	113	97
2. Non-Social and Human Disciplines	476	291	185
Medicine	155	112	43
Architecture	151	88	63
Agriculture	170	91	79
All Subgroups	1069	684	385

3.3. Instruments

The IAUEPT is a significant assessment tool applied to make important decisions regarding PhD candidates. It is important to assess and address DIF and DDF in this test in order to make sure that the test items are valid and impartial for all groups of testees. For the purpose of analyzing the IAUEPT, the SPSS program version 22 was utilized. The mean score for the entire test was found to be 78.33. The Cronbach's alpha reliability coefficient for the entire assessment was calculated to be .84 indicating a high level of consistency. Table 2 provides in-depth data on the descriptive statistics and reliabilities for every part of the

IAUEPT along with the number of items for each section.

Table 2.

Approximations of Reliability for the IAUEPT

Test	No. Items	Mean	SD	Variance	Min.	Max.	Range	Reliability
Vocabulary	25	21.6	2.1	4.41	15	25	10	.44
Structure I	25	19.6	3.1	9.61	1	25	24	.68
Structure II	15	8.6	2	4	2	14	12	.26
Reading	35	28.5	6.6	43.56	0	35	35	.92
Total	100	78.33	8.88	78.87	45	94	49	.84

In this study, the evaluation of DIF and DDF across academic field was conducted with the help of the Rasch model (Rasch, 1960/1980). Additionally, the estimation of the model was performed applying the Winsteps computer program (Linacre, 2009).

3.4. Data Collection Procedure

Given that the test takers' response sheets were located at the central organization of Islamic Azad University in Tehran, the administrators of Najafabad branch successfully obtained a substantial number of these response sheets through coordinating with the headmasters of the central exam administration. Furthermore, the central organization was given an assurance that the information contained in the answer sheets would not be disclosed. The segregation of the answer sheets based on the academic fields and other criteria facilitated the insertion of essential data, including the testees' academic fields, test scores, and correct choices into a Microsoft Excel file to be read by the Winsteps software.

DIF analysis is a comparative method based on contrasting the performance of two groups. Conducting DIF across more than two groups complicates the analysis, making it too sophisticated to interpret the results when multiple group comparisons are involved. Therefore, the 1069 examinees were divided into two academic field groups named Social and Human Disciplines (SHD) and Non-Social and Human Disciplines (N-SHD) for the DIF and DDF analyses. The SHD group included Humanities, Psychology, and Educational Sciences academic fields, while the N-SHD group comprised Medicine, Architecture, and Agriculture academic fields.

3.5. Data Analysis Procedure

It is important to highlight that DIF analysis does not require sampling. Sampling becomes necessary when new data needs to be collected to ensure representativeness of the entire population. In this study, it was crucial to analyze the responses of the entire population to the test items in order to detect academic field-based DIF and DDF. The DDF analysis was then performed separately for Choices A, B, C, and D across academic field, following three phases:

1. Choice A = 1 and other choices = 0 were rescored for the IAUEPT.
2. A standard DIF was carried out across different fields of study.
3. Items that exhibit DIF were recognized. If an item demonstrates DIF for the sake of the N-SHD group, it signifies that Choice A has attracted a greater number of participants from the N-SHD group than those of the SHD group. DDF and DIF are the same in this item if Choice A is selected accurately, and DDF is reduced to DIF.

In general, the above-named steps were repeated to explore other choices, and the Rasch model was utilized during the data analysis process.

Additionally, The Winsteps software was utilized to calculate the Rasch model. The procedure for estimating the Rasch model employing the Winsteps software involves the following steps:

1. The data was prepared in a specific format to be read by Winsteps. This included a file with the responses of each testee to each item along with any demographic information about academic field.

2. The items in the test, the number of response categories for each item, and any constraints on the model parameters were specified.

3. An iterative algorithm to estimate the Rasch model parameters was applied. The program started with initial estimates of each testee's ability and each item's difficulty, and then adjusted these estimates based on the responses of the testees to the items.

4. After the analysis was completed, the Winsteps provided several measures of model fit, including the overall fit of the model to the data and individual item fit statistics. These measures were applied to assess the quality of the Rasch model fit to the data.

5. After evaluating model fit, the results of the analysis were interpreted. This included examining estimates of each testee's ability and each item's difficulty, as well as the existence of academic field-based DIF and DDF.

4. Results

4.1. Measurement and Fit Values for the IAUEPT Items.

Table 3 presents the item statistics and fit values. The interpretation is based on the information provided in this Table (Appendix 1). The column tagged 'measure' in the Table discloses the degree of difficulty per item. A greater measure demonstrates a more challenging item. As reported by Table 3, Item 55 is the extremely challenging item, while Item 25 is the simplest. The column marked 'S.E.' discloses the standard error of the item measures. A lower S.E. demonstrates a truer estimation of item difficulties. The values of infit and outfit mean square provide information about the fit of the items to the Rasch model (Baghaei et al., 2017). Values underneath 1.30 are considered satisfactory. However, as can be seen in Table 3, Items 15, 54, 61, and 63 have outfit mean square values exceeding 1.30, indicating that they do not fit the Rasch model. The column labeled 'PT-Measure CORR.' exhibits the point-biserial correlation between the items and person ability measures. Higher point-measure correlations exhibit better item discrimination. Values above .20 are considered satisfactory (Linacre, 2023). Based on the point-measure correlations, it is absolutely clear that many items in the test have inferior property and have to be eradicated.

4.2. Dealing with the First Research Question

The initial question of the current study sought to examine the significant impact of the participants' academic field on the presence of DIF in the IAUEPT. The findings pertaining to this question are indicated in the following Table with only the items demonstrating academic field-related DIF being presented for the sake of convenience.

Table 4.

Differential Item Functioning Statistics Across Academic Field

Item Number	Person Class	DIF Measure	Person Class	DIF Measure	DIF Contrast	Rasch-Welch		
						t	df	Prob
I76	SHD	.91	N-SHD	-.15	1.06	3.00	204	.0030
I80	SHD	.81	N-SHD	.07	.75	2.18	204	.0305

Table 4 presents the pertinent statistics regarding DIF across the testees' academic fields. The column labeled 'DIF Measure' on the left showcases the level of difficulty for the items within the SHD group, while the corresponding column on the right represents the item difficulty for the N-SHD group. 'Contrast' refers to the disparity between two measures of difficulty. This disparity is examined for statistical significance applying a t-test. The 't' column presents the *t* statistic value, the 'd.f.' column indicates the degrees of freedom, and the 'Prob.' column displays the p-value.

The null hypothesis posits that the academic field of the participants does not bring about significant DIF in the IAUEPT. Conversely, the alternate hypothesis suggests that testees' field of study does lead to significant DIF in the IAUEPT. The findings presented in Table 4 imply that the academic field of the participants did indeed result in significant DIF in the IAUEPT. This is evident from the p-values which are less than .05 demonstrating that the difficulty difference between the SHD and N-SHD groups for an item is statistically significant, and the item exhibits DIF across the examinees' academic fields. The 'Prob.' column in Table 4 further manifests that Items 76 and 80 exhibit academic field-related DIF with both items favoring the N-SHD group. It is worth noting that a smaller DIF measure indicates an easier item. Accordingly, the results confirm that the attendees' major led to significant DIF in the IAUEPT, bringing about the rejection of the null hypothesis.

4.2.1. Visual Representation of DIF Across Academic Field

Figure 1.

SHD and N-SHD Item Characteristic Curves for Item 76

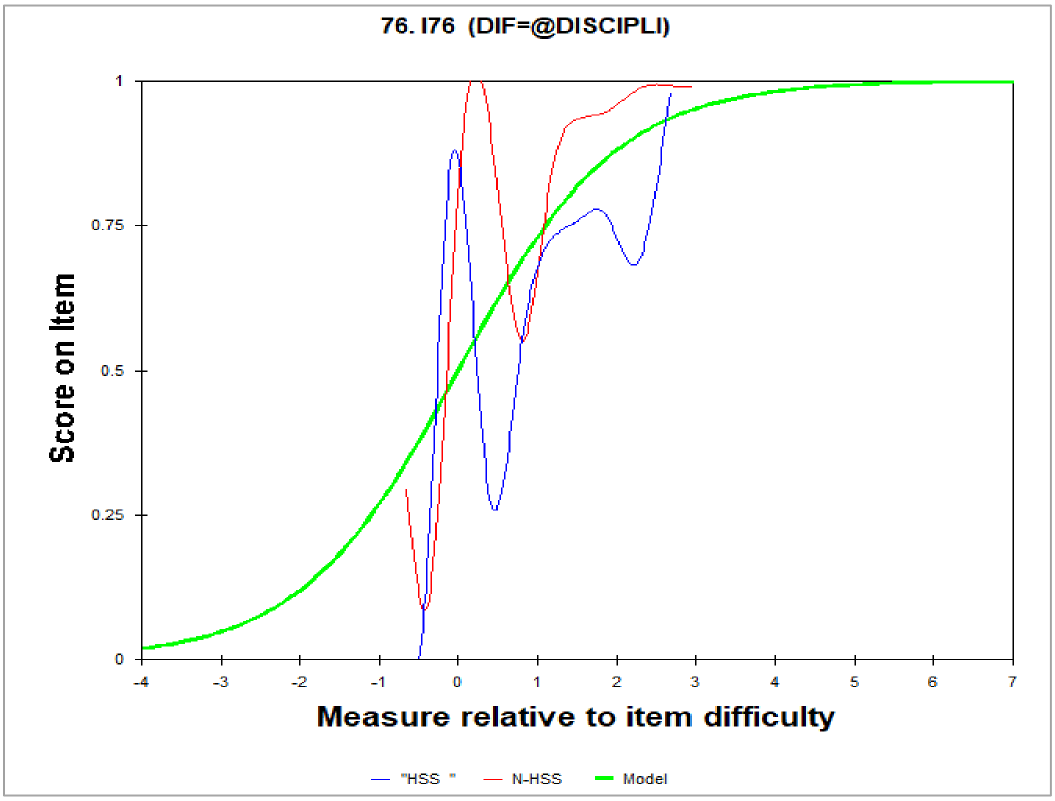

Figure 1 illustrates the pairwise Item Characteristic Curves (ICCs) for Item 76 for the SHD and N-SHD groups. The x-axis signifies person ability, while the axis of y designates the likelihood of an accurate answer. The standard Rasch model is depicted by the S-shaped ICC. For an item not to have DIF, the ICCs across the two subsamples (i.e. SHD and N-SHD) should overlap. According to this figure, the two ICCs do not overlap at most parts which is evidence that the item has academic field DIF. In other words, the item is not equally difficult for the two groups, and hence exhibits DIF.

4.3. Dealing with the Second Research Question

The second research question that needed to be investigated aimed to determine if the attendees' major has a notable impact on the degree of DDF in the IAUEPT. In order to accomplish this objective, the DDF was assessed for each choice individually. The evaluation of DDF was conducted in three stages as outlined in the procedure section. Note that, because of space restrictions, merely DDF items across testees' fields of study are displayed in the following Tables.

4.3.1. Academic Field-based Differences in DDF for Choice A

Table 5 displays the results of the DDF analysis for Choice A across field of study. The column labeled 'DIF Measure' on the left side shows the degree to which the SHD group has chosen 'Choice A', while 'DIF Measure' on the right side displays the degree to which the N-SHD group has decided on 'Choice A'. A lesser DIF measure indicates an upper level of acceptance of the choice within a particular group.

Table 5.

Differential Distractor Functioning for Choice A Across Academic Field

Item Number	Person Class	DIF Measure	Person Class	DIF Measure	DIF Contrast	Rasch-Welch		
						t	df	Prob
I79	SHD	.53	N-SHD	2.22	-1.69	-2.52	187	.0127

The results presented in Table 5 indicate that merely Item 79 has a p-value less than .05, suggesting that Choice A in this item exhibits DDF across examinees' academic fields. Specifically, the DIF measure of this item is .53 for the SHD group and 2.22 for the N-SHD group. This shows that Choice A was more preferred by the SHD group. Furthermore, the dissimilarity in the DIF measure across the two groups (DIF Contrast=-1.69) is statistically significant, as evidenced by a t-value of -2.52 with 187 degrees of freedom and a p-value of .0127.

4.3.2. Academic Field-based Differences in DDF for Choice B

Table 6 outlines the consequences of the DDF analysis for Choice B across attendees’ field of study. The Prob. column indicates that Items 2, 44, and 76 have p-values below the threshold of .05, suggesting that Choice B exhibits DDF across academic fields in these items. Specifically, Item 2 demonstrates a DIF measure of 1.25 for the SHD group and 3.66 for the N-SHD group, indicating that Choice B in this item is more preferred by the SHD group. The variance in the DIF measure across the two groups (DIF Contrast=-2.42) is statistically significant (*t*=-2.28, *df*=181, *p*=.0239). Similarly, Item 44 exhibits a DIF measure of 2.53 for the SHD group and 0.36 for the N-SHD group, indicating a preference for Choice B among the N-SHD group. The variance in the DIF measure across the two groups (DIF Contrast=2.16) is statistically significant (*t*=3.12, *df*=163, *p*=.0022). Lastly, Item 76 demonstrates a DIF measure of -2.54 for the SHD group and -3.73 for the N-SHD group, indicating a stronger preference for Choice B amongst the N-SHD group. The variance in the DIF measure across the two groups (DIF Contrast=1.19) is statistically significant (*t*=3.28, *df*=196, *p*=.0012).

Table 6.

Differential Distractor Functioning for Choice B Across Academic Field

Item Number	Person Class	DIF Measure	Person Class	DIF Measure	DIF Contrast	Rasch-Welch		
						t	df	Prob
I2	SHD	1.25	N-SHD	3.66	-2.42	-2.28	181	.0239
I44	SHD	2.53	N-SHD	.36	2.16	3.12	163	.0022
I76	SHD	-2.54	N-SHD	-3.73	1.19	3.28	196	.0012

4.3.3. Academic Field-based Differences in DDF for Choice C

Table 7 presents the DDF consequences for Choice C across the examinees’ major. The Prob. column in this Table signifies that Items 3, 48, 76, and 80 have p-values less than the threshold of .05. This suggests that Choice C in these items exhibits DDF across academic fields. Specifically, Item 3 shows a DIF measure of 2.93 for the SHD group and 1.03 for the N-SHD group. This signifies that Choice C in Item 3 is more preferred by the N-SHD group. The dissimilarity in the DIF measure across the two groups (DIF Contrast=1.90) is statistically significant, as evidenced by a t-value of 2.39 with 175 degrees of freedom (df) and a p-value of .0180. Similarly, Item 48 manifests a DIF measure of .23 for the SHD group and 1.17 for the N-SHD group. This suggests that Choice C in this item is more chosen by the SHD group. The dissimilarity in the DIF measure across the two groups (DIF Contrast=-.94) is statistically significant, with a t-value of -2.01 and 201 df, yielding a p-value of .0461. Furthermore, Item 76 exhibits a DIF measure of -.58 for the SHD group and .54 for the N-SHD group. It can be inferred Choice C in this item is more chosen by the SHD group. The dissimilarity in the DIF measure between the two groups (DIF Contrast=-1.11) is statistically significant, with a t-value of -2.98 and 196 df, resulting in a p-value of .0033. Lastly, Item 80 demonstrates a DIF measure of -.41 for the SHD group and .64 for the N-SHD group. It can be stated that Choice C in this item is greatly preferred by the SHD group. The dissimilarity in the DIF measure between the two groups (DIF Contrast=-1.05) is statistically significant (*t*=-2.70, *df*=196, *p*=.0075).

Table 7.

Differential Distractor Functioning for Choice C Across Academic Field

Item Number	Person Class	DIF Measure	Person Class	DIF Measure	DIF Contrast	Rasch-Welch		
						t	df	Prob
I3	SHD	2.93	N-SHD	1.03	1.90	2.39	175	.0180
I48	SHD	.23	N-SHD	1.17	-.94	-2.01	201	.0461
I76	SHD	-.58	N-SHD	.54	-1.11	-2.98	196	.0033
I80	SHD	-.41	N-SHD	.64	-1.05	-2.70	196	.0075

4.3.4. Academic Field-based Differences in DDF for Choice D

Table 8 contains the outcomes of the DDF analysis for Choice D across academic fields. The Prob. column in Table 8 designates that Items 26, 44, and 80 have p-values lesser than .05, suggesting that Choice D exhibits DDF across academic fields in these items. Item 26 illustrates a DIF measure of .11 for the SHD group and 1.10 for the N-SHD group. This proves that Choice D was more accepted among the SHD group. The dissimilarity in the DIF measure across the two groups (DIF Contrast=-.99) is statistically significant, as evidenced by a t-value of -2.22 with 200 degrees of freedom (df) and a p-value of .0273. Likewise, Item 44 displays a DIF measure of -3.97 for the SHD group and -3.07 for the N-SHD group, indicating a preference for Choice D among the SHD group. The variance in the DIF measure across the two groups (DIF Contrast=-.89) is statistically significant, with a t-value of -2.03, 185 df, and a p-value of .0442. Finally, Item 80 exhibits a DIF measure of -2.41 for the SHD group and -3.26 for the N-SHD group, suggesting a preference for Choice D among the N-SHD group. The variance in the DIF measure across the two groups (DIF Contrast=.85) is statistically significant (*t*=2.40, *df*=197, *p*=.0172).

Table 8.

Differential Distractor Functioning for Choice D Across Academic Field

Item Number	Person Class	DIF Measure	Person Class	DIF Measure	DIF Contrast	Rasch-Welch		
						t	df	Prob
I26	SHD	.11	N-SHD	1.10	-.99	-2.22	200	.0273
I44	SHD	-3.97	N-SHD	-3.07	-.89	-2.03	185	.0442
I80	SHD	-2.41	N-SHD	-3.26	.85	2.40	197	.0172

The null hypothesis posits that the academic field of the participants does not bring about significant DDF in the IAUEPT. Conversely, the alternate hypothesis suggests that the attendees' major does indeed lead to significant DDF in the IAUEPT. The findings presented in Tables 5, 6, 7, and 8 demonstrate that the items examined throughout these Tables experience DDF across the candidates' academic fields. Consequently, it can be concluded that the participants' academic field has a significant impact on the occurrence of DDF in the IAUEPT, leading to the rejection of the null hypothesis.

5. Discussion

The outcomes of DIF evaluation across field of study illustrated that Item 80 exhibited academic field-related DIF conducive to the N-SHD group. This item was derived from a passage that was titled "*Methods of Waste Management: Pros and Cons*".

Item 80. The main purpose of the passage is ………….…..... .

A) to persuade readers that recycling is a waste of resources.

B) to compare and contrast recycling and landfilling.

C) to convince readers to recycle and compost.

D) to inform readers of methods of waste management

One credible clarification for why this item was in the interest of the N-SHD group is that the phrase "*waste management"* in Choice D, the correct choice, is directly related to the medicine and agriculture academic fields that belong to the N-SHD group. The field of waste management and recycling is primarily focused on by researchers in the natural sciences. However, there is a growing number of scholars who are approaching the topic from a humanistic and societal standpoint. This trend indicates an increasing interest in understanding waste management beyond its technical aspects. Students in medicine and agriculture often encounter waste management practices in their practical training or internships. Understanding different methods of waste management is essential for them to implement proper procedures in their fields of study. In medicine, proper waste management is crucial to prevent the spread of infections and diseases in healthcare settings (Kenny & Priyadarshini, 2021). Additionally, in agriculture, waste management plays a role in sustainable farming practices and environmental conservation (Obi, Ugwuishiwu, & Nwakaire, 2016). Seemingly, the N-SHD group accepted Choice D due to the fact that the phrase "*waste management"* was related to their fields of study as well as their areas of interest.

Item 44. Jim …………… trying to pass his driving test, but he fails every time.

A) would keep B)would have kept

C) was kept D)keeps

The outcomes for Choice B across academic field-based DDF revealed that Choice B in Item 44 was more approved amongst the N-SHD group. The structure section was where this item originated. A number of studies have observed that the students in the fields of social and human disciplines demonstrate satisfactory proficiency in grammar assessments. For example, Zand-Moghadam et al. (2018) stated that EAP courses of social and human disciplines pay the deserved attention to the sub-skills of syntax and structure. There are multiple factors that may contribute to the relatively satisfactory performance of students in the social and human disciplines on grammar assessments. One potential explanation is that academic fields within the humanities and social sciences frequently necessitate proficient reading, writing, and communication abilities. Therefore, students pursuing these fields are exposed to a diverse array of texts, which can contribute to the enhancement of their grammatical proficiency. Another possible rationale is that courses in the social and human disciplines frequently entail the application of critical analysis and interpretation to textual materials. This cognitive process necessitates meticulous attention to grammatical and syntactical elements as students grapple with intricate ideas and concepts (Kennedy & Kennedy, 2012; Luckett, 2016). Apparently, Choice B, the distractor, was more desired among the N-SHD group since it manifested greater syntactic complexity compared to the other choices, and the group strived to find the correct choice via guessing because of their insufficient knowledge of structure.

Item 3. A(n) ……………. referee does not favor one team in relation to another.

A) intimate B) biased

C) subjective D) impartial

The findings from the examination of the DDF based on the candidates' academic fields evidenced that Choice C in Item 3, derived from the vocabulary section, was more accepted amongst the N-SHD group which comprises the medicine and architecture academic fields. It is worth noting that Choice D is the correct choice. One possible explanation for the confusion among students regarding the distinction between impartiality and subjectivity or bias could be attributed to their limited comprehension of the concept of objectivity or neutrality within a specific context (Baron, 2000; Kunda, 1990).

It appears that examinees majoring in medicine displayed a preference for Choice C over Choice D potentially due to their limited exposure to vocabulary and concepts pertaining to objectivity and impartiality within their field of study. In the realm of medicine, students may be more accustomed to subjective symptoms or judgments derived from their personal experiences or perceptions (Croskerry, 2003).

There are multiple factors that may explain why the testees majoring in architecture exhibited a preference for the term "*subjective*" in Choice C. Architecture as an academic field inherently involves individual preferences in relation to design and aesthetics. Students studying architecture are frequently encouraged to cultivate their own distinct style and approach to design, which inherently entails subjective elements. Consequently, their inclination towards selecting the term "*subjective*" may be attributed to its alignment with their comprehension of the creative process (Balık, 2016; Koch, 2018). Furthermore, architecture students typically prioritize the enhancement of their technical expertise and understanding in areas such as design principles, construction techniques, and architectural history. However, they may not receive comprehensive instruction in language proficiency or be exposed to specialized vocabulary pertaining to the language usage (Groat & Ahrentzen, 1996; Swales et al., 2001). Hence, it seems that this limited exposure may potentially hinder their ability to differentiate between terms such as partiality and subjectivity.

5.1. Justifications for Different Test Performances Observed Among Test Takers Majoring in Different Academic Fields

Students from different fields of study receive different grades in language assessments. Today, this issue has attracted the attention of many researchers and linguists. This variation in performance can be attributed to several reasons, some of which have been discussed in this study. Some fields of study are closely related to language. Consequently, students are more exposed to vocabulary and language structure, leading them to subconsciously acquire many language structures. Another factor is related to a person's familiarity with language learning strategies. Research has shown that, compared to students in other fields, science students are more familiar with specialized vocabulary and writing strategies. This familiarity enables them to perform better on language assessments. Additionally, an individual's level of motivation and interest in learning a second language can result in a higher score on the language assessment. In some fields, students need to be sufficiently familiar with a second language. This familiarity enhances students' interest in the language and improves their performance in language assessments compared to students in other fields. Another factor that can positively impact students' performance in language competency assessments is intellectual abilities, such as attentiveness and processing speed. Students in fields of study that require higher intellectual abilities, such as philosophy or math, are able to score higher on language competency assessments by reason of their enhanced intellectual abilities (Daneman & Merikle, 1996).

6. Conclusion

The objective of the current study was to appraise DIF and DDF on the IAUEPT via examining academic fields of PhD candidates. The results of DIF analysis demonstrated that two items were influenced by DIF based on the attendees' major. These items were in the interest of the N-SHD group. Besides, the consequences for DDF assessment determined a series of problematic items. Hence, Choice A revealed one item, Choice B had three items, Choice C consisted of four items, and Choice D contained three items. It is worth mentioning that seven choices were more accepted by the SHD group, whereas four choices were better preferred amidst the N-SHD group.

Different sections of the IAUEPT were also examined. According to the outcomes of DIF, the items in the reading comprehension section were more favored by the attendees in the N-SHD group. Based on the outcomes of DDF, both groups were interested in the choices of the reading comprehension and vocabulary sections. However, the SHD group preferred the choices in the structure subdivision greater than the other group. In summary, the outcomes of the current study are very useful for assessment analysts who aim to ensure the validity of high-stakes assessments.

The implications of this study are of great value to those who are involved in assessment development, education, and policy-making. Firstly, testees can benefit from this study by gaining awareness of potential biases in assessments that are specific to their academic fields. This knowledge can assist them in better preparing for the assessment by focusing on areas that may pose greater challenges due to DIF or DDF. Additionally, the findings have significant implications for teaching strategies. Through recognizing academic field-specific DIF, educators can gain insight into potential differences in how students from different fields of study comprehend and respond to certain assessment items. This understanding can enable educators to tailor their teaching strategies to ensure that all students have equal opportunities to learn and succeed. Overall, further analytical studies dedicated to investigating academic field gaps in language assessments may be necessary in order to identify the sources of academic field-specific DIF and DDF.

Every academic research can come across some limitations and the present study is not an exception. In the present study, only one method was utilized for each analysis of DIF and DDF, which may be considered a limitation. Furthermore, the study solely focused on examining candidates' academic fields as a potential source of bias in a high-stakes multiple-choice assessment. Future investigations could expand the scope by exploring other variables such as race, ethnicity, or socioeconomic status. Another limitation pertains to the assessment format and content. The use of a multiple-choice assessment format may not fully capture all dimensions of knowledge and skills relevant to the examinees' fields of study, potentially impacting the assessment results. Additionally, the specific items included in the assessment may not comprehensively assess all aspects of knowledge and skills associated with each academic field, thereby limiting the accuracy of the DIF and DDF analyses. In sum, while this study offers valuable insights into DIF and DDF on the IAUEPT, it is important to consider these limitations when interpreting the findings.

Based on the aforementioned limitations, it is suggested that further research should be conducted to address these issues. For instance, more detailed studies could be undertaken to investigate the influence of assessment format, such as computer-based versus paper-based, on the results of DIF and DDF. In relation to assessment format, it would be beneficial for researchers to compare the outcomes of multiple-choice examinations with other types of assessment. This comparative analysis would help identify any particular biases clearly related to specific assessment designs and give a glimpse into strategies for mitigating them. Additionally, it is recommended that future research examine the persistence or changes in DIF and DDF after a period of time. Repeating the investigation during distinct periods will boost the validity of the assessments and contribute to a more wide-ranging perception of their generalizability in diverse situations.

References

Abedi, J., Leon, S., & Kao, J. C. (2008). Examining differential item functioning in reading assessments for students with disabilities. *CRESST Report*, *744*. https://doi.org/10.1037/E642932011-001

Adibatmaz, F. B. K., & Yildiz, H. (2020). The effects of distractors to differential item functioning in Peabody picture vocabulary test. *Journal of Theoretical Educational Science*, *13*(3), 530-547. https://doi.org/10.30831/akukeg.622180

Ahmadi, A., & Jalili, T. (2014). A confirmatory study of differential item functioning on EFL reading comprehension. *Applied Research on English Language*, *3*(2), 55-68. https://doi:10.22108/are. 2014.15489

Ajideh, P., Yaghoubi-Notash, M., & Babaee Bormanaki, H. (2022). Native language-based DIF across the subtests: A study of the Iranian national university entrance exam. *Journal of English Language Teaching and Learning University of Tabriz*, *14*(30), 39-56. https://doi: 10.22034/elt.2022.51852.2491

Alavi, S. M., & Bordbar, S. (2018). Differential item functioning analysis of high-stakes test in terms of gender: A Rasch model approach. *MOJES: Malaysian Online Journal of Educational Sciences*, *5*(1), 10-24. https://mojes.um.edu.my/article/view/12631

Alavi, S., Ali Rezaee, A., Amirian, S. (2012). Academic discipline DIF in an English language proficiency test. *Journal of English Language Teaching and Learning University of Tabriz*, *3*(7), 39-65.

Ayoobiyan, H., & Ahmadi, A. (2023). Detecting halo effects across rubric criteria in L2 writing assessment: A many-facet Rasch analysis. *Applied Research on English Language*, *12*(1), 159-176. https://doi: 10.22108/are.2022.132503.1848

Baghaei, P. (2021). *Mokken scale analysis in language assessment.* Münster, Germany: Waxmann Verlag.

Balık, D. (2016). Ornament: The Politics of architecture and subjectivity. *The Journal of Architecture*, *21*(8), 1336–1339. https://doi.org/10.1080/13602365.2016.1257277

Barati, H., Ketabi, S., & Ahmadi, A. (2006). Differential item functioning in high-stakes tests: The effect of field of study. *Iranian Journal of Applied Linguistics*, *19*(2), 27-42.

Baron, J. (2000). *Thinking and deciding* (3rd ed.). Cambridge University Press

Bowles, M. A. (2022). Using instructor judgment, learner corpora, and DIF to develop a placement test for Spanish L2 and heritage learners. *Language Testing*, *39*(3), 355-376. https://doi.org/10.1177/02655322221076033

Camilli, G., & Penfield, D. A. (1997). Variance estimation for differential test functioning based on the Mantel-Haenszel statistics. *Journal of Educational Measurement*, *34*(2), 123–139. https://doi.org/10.1111/j.1745-3984.1997.tb00510.x

Clauser, B. E., & Mazor, K. M. (2005). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, *17*(1), 31–44. https://doi.org/10.1111/j.1745-3992.1998.tb00619.x

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*, *78*(8), 775-780. https://doi.org/10.1097/00001888-200308000-00003

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422-433. https://doi.org/10.3758/BF03214546

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research*, *87*(6), 1082–1116. http://www.jstor.org/stable/44667687

Green, B. F., Crone, C. R., & Folk, V. G. (1989). A method for studying differential distractor functioning. *Journal of Educational Measurement*, *26*(2), 147–160. https://doi.org/10.1111/j.1745-3984.1989.tb00325.x

Groat, L. N., & Ahrentzen, S. (1996). Reconceptualizing architectural education for a more diverse future: Perceptions and visions of architectural students. *Journal of Architectural Education*, *49*(3), 166-183. https://doi.org/10.1080/10464883.1996.10734679

Hoshino, Y. (2013). Relationship between types of distractor and difficulty of multiple-choice vocabulary tests in sentential context. *Language Testing in Asia*, *3*(1), 1-14. https://doi.org/10.1186/2229-0443-3-16

Jafaripour, S., Tabatabaei, O., Salehi, H., & Vahid Dastjerdi, H. (2024). Applying IRT Model to Determine Gender and Discipline-based DIF and DDF: A Study of the IAU English Proficiency Test. *International Journal of Language Testing*, *14*(1), 56-74. https://doi.org/10.22034/ijlt.2023.407117.1268

Karami, H. (2012). An introduction to differential item functioning. *The International Journal of Educational and Psychological Assessment*, *11*(2), 59-76.

Kato, K., Moen, R. E., & Thurlow, M. L. (2009). Differentials of a state reading assessment: Item functioning, distractor functioning, and omission frequency for disability categories. *Educational Measurement: Issues and Practice*, *28*(2), 28-40. https://doi.org/10.1111/j.1745-3992.2009.00145.x

Kennedy, M. L., & Kennedy, W. J. (2012). *Writing in the disciplines: A reader and rhetoric for academic writers*. Pearson.

Kenny, C., & Priyadarshini, A. (2021). Review of current healthcare waste management methods and their effect on global health. *Healthcare*, *9*(3), 284. https://doi.org/10.3390/healthcare9030284

Khalaf, M. A., & Omara, E. M. N. (2022). Rasch analysis and differential item functioning of English language anxiety scale (ELAS) across sex in Egyptian context. *BMC psychology*, *10*(1), 242. https://doi.org/10.1186/s40359-022-00955-w

Kim, S.-H., & Cohen, A. S. (1995). A comparison of Lord's chi-square, Raju's area measures, and the Likelihood Ratio Test on detection of differential item functioning. *Applied Measurement in Education*, *8*(4), 291–312.

Koch, D. (2018). On architectural space and modes of subjectivity: Producing the material conditions for creative-productive activity. *Urban Planning*, *3*(3), 70-82. https://doi.org/10.17645/up.v3i3.1379

Koon, S., & Kamata, A. (2013). An applied examination of methods for detecting differential distractor functioning. *International Journal of Quantitative Research in Education*, *1*(4), 364-382. https://doi.org/10.1504/IJQRE.2013.058306

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Linacre, J. M. (2009). *WINSTEPS Rasch Measurement* (Version 3.73) [Computer software]. Chicago, IL:

Winsteps.com.

Linacre, J. M. (2023b). Winsteps® Rasch measurement computer program User's Guide. Version 5.6.0. Portland, Oregon: Winsteps.com.

Luckett, K. (2016). Making the implicit explicit: The grammar of inferential reasoning in the humanities and social sciences. *Universal Journal of Educational Research*, *4*(5), 1003-1015. https://doi.org/10.13189/ujer.2016.040510

Mapuranga, R., Dorans, N. J., & Middleton, K. (2008). A review of recent developments in differential item functioning. *ETS Research Report Series*, *2008*(2), i–32. https://doi.org/10.1002/j.2333-8504.2008.tb02129.x

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5-11. https://doi.org/10.3102/0013189X018002005

Moradi, E., Ghabanchi, Z., & Pishghadam, R. (2022). Reading comprehension test fairness across gender and mode of learning: Insights from IRT-based differential item functioning analysis. *Language Testing in Asia*, *12*(1), 1-18. https://doi.org/10.1186/s40468-022-00192-3

Obi, F., Ugwuishiwu, B., & Nwakaire, J. (2016). Agricultural waste concept, generation, utilization and management. *Nigerian Journal of Technology*, *35*(4), 957. https://doi.org/10.4314/njt.v35i4.34

Penfield, R. D. (2008). An odds ratio approach for assessing differential distractor functioning effects under the nominal response model. *Journal of Educational Measurement*, *45*(3), 247–269. http://www.jstor.org/stable/20461895

Penfield, R. D. (2010). Modeling DIF effects using distractor-level invariance effects: Implications for understanding the causes of DIF. *Applied Psychological Measurement*, *34*(3), 151-165. https://doi.org/10.1177/0146621609359284

Penfield, R. D. (2011). How are the form and magnitude of DIF effects in multiple-choice items determined by distractor-level invariance effects? *Educational and Psychological Measurement*, *71*(1), 54–67. https://doi.org/10.1177/0013164410387340

Penfield, R. D., & Lam, T. C. (2000). Assessing differential item functioning in performance assessment: Review and recommendations. *Educational Measurement: Issues and Practice*, *19*(3), 5-15. https://doi.org/10.1111/j.1745-3992.2000.tb00033.x

Penton, H., Dayson, C., Hulme, C., & Young, T. (2022). An investigation of age-related differential item functioning in the EQ-5D-5L using item response theory and logistic regression. *Value in Health*, *25*(9), 1566-1574. https://doi.org/10.1016/j.jval.2022.03.009

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research, 1960. (Expanded edition, Chicago: The University of Chicago Press, 1980). https://doi.org/10.7208/chicago/9780226702764.001.0001

Sadeghi, K., & Abolfazli Khonbi, Z. (2017). An overview of differential item functioning in multistage computer adaptive testing using three-parameter logistic item response theory. *Language Testing in Asia*, *7*(1), 1-16. https://doi.org/10.1186/s40468-017-0038-z

Sazegar, Z., Ashraf, H., & Motallebzadeh, K. (2021). Constructing and validating an EFL hidden curriculum scale using the Rasch model. *Applied Research on English Language*, *10*(1), 1-32. https://doi: 10.22108/are.2020.121574.1540

Stemler, S. E., & Naples, A. (2021). Rasch measurement v. item response theory: Knowing when to cross the line. *Practical Assessment, Research and Evaluation*, *26*(1), 11. https://doi.org/10.7275/v2gd-4441

Suh, Y., & Talley, A. E. (2015). An empirical comparison of DDF detection methods for understanding the causes of DIF in multiple-choice items. *Applied Measurement in Education*, *28*(1), 48-67. https://doi.org/10.1080/08957347.2014.973560

Swales, J. M., Barks, D., Ostermann, A. C., & Simpson, R. C. (2001). Between critique and accommodation: Reflections on an EAP course for masters of architecture students. *English for Specific Purposes*, *20*(1), 439–458. https://doi.org/10.1016/S0889-4906(01)00020-5

Swamanithan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, *27*(4), 361-370. http://www.jstor.org/stable/1434855

Takala, S., & Kaftandjieva, F. (2000). Test fairness: A DIF analysis of an L2 vocabulary test. *Language Testing*, *17*(3), 323-340. https://doi.org/10.1177/026553220001700303

Terao, T., & Ishii, H. (2020). A comparison of distractor selection among proficiency levels in reading tests: A focus on summarization processes in Japanese EFL learners. *SAGE Open*, *10*(1), 1–14. https://doi.org/10.1177/2158244020902087

Törmäkangas, K. (2011). Advantages of the Rasch measurement model in analysing educational tests: An applicator's reflection. *Educational Research and Evaluation*, *17*(5), 307-320. https://doi.org/10.1080/13803611.2011.630562

Tsaousis, I., Sideridis, G., & Al-Saawi, F. (2018). Differential distractor functioning as a method for explaining DIF: The case of a national admissions test in Saudi Arabia. *International Journal of Testing*, *18*(1), 1–26. https://doi.org/10.1080/15305058.2017.1345914

Zand-Moghadam, A., Meihami, H., & Ghiasvand, F. (2018). Exploring the English language needs of EAP students of humanities and social sciences in Iran: A triangulated approach. *Issues in Language Teaching*, *7*(1), 135-164. https://doi.org/10.22054/ilt.2019.47351.434

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (ordinal) item scores*. Ottawa, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zumbo, B. D. (2007). Three generations of DIF analysis: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly, 4(2)*, 223-233. https://doi.org/10.1080/15434300701375832

Appendix 1

Table 3.

Measurement and Fit Values for the IAUEPT Items.

Item	Measure	SE	Infit MNSQ	Outfit MNSQ	PT-Measure
1	-.47	.22	1.14	1.24	.02
2	-1.23	.29	1.03	.93	.16
3	.51	.17	1.01	.96	.29
4	-.38	.21	1.09	1.04	.12
5	-.68	.24	1.04	.99	.18
6	.05	.19	1.09	1.09	.13
7	-.80	.25	1.07	1.05	.12
8	.31	.17	1.00	.96	.29
9	.65	.16	1.03	1.03	.23
10	-5.10	1.82	MINIMUM MEASURE		.00
11	-2.78	.58	.96	.40	.24
12	-.29	.21	1.07	.96	.18
13	.02	.19	1.05	1.14	.17
14	-2.06	.42	.97	.63	.23
15	-.80	.25	1.14	1.41	-.04
16	-1.07	.27	1.05	1.09	.11
17	-.52	.22	1.07	1.18	.11
18	1.13	.15	1.02	1.03	.25
19	-2.06	.42	1.02	1.19	.07
20	-.43	.22	1.07	1.03	.15
21	-1.32	.30	1.08	1.26	.02
22	-1.07	.27	.99	.85	.24
23	-1.23	.29	1.09	1.35	.01
24	2.07	.15	1.16	1.23	.02
25	-5.10	1.82	MINIMUM MEASURE		.00
26	.46	.17	1.14	1.15	.07
27	2.45	.15	.99	1.08	.25
28	-1.32	.30	1.00	.91	.20
29	-.13	.20	1.09	1.04	.15
30	.31	.17	1.04	1.03	.22
31	-1.41	.32	.97	.76	.25
32	-.25	.20	1.01	.88	.27
33	-2.78	.58	.97	.42	.23
34	-.02	.19	1.07	1.03	.18
35	.75	.16	1.10	1.09	.15
36	2.33	.15	1.02	1.09	.21
37	-.74	.24	1.03	1.21	.14
38	-.57	.23	1.07	1.13	.11
39	1.88	.15	1.09	1.16	.12
40	-.05	.19	1.09	1.07	.13
41	-.57	.23	1.05	1.10	.14

42	-1.63	.35	.97	.77	.24
43	-.43	.22	1.10	1.23	.06
44	-.05	.19	1.11	1.12	.10
45	2.67	.16	.96	1.04	.29
46	-1.90	.39	.99	.84	.18
47	-.05	.19	1.06	1.05	.17
48	-.34	.21	1.00	.85	.28
49	.09	.18	1.07	1.02	.18
50	-.74	.24	1.05	.99	.16
51	.09	.18	1.13	1.17	.07
52	.54	.17	1.13	1.11	.11
53	1.83	.15	1.03	1.09	.22
54	2.43	.15	1.16	1.35	-.02
55	3.00	.17	1.05	1.16	.13
56	.40	.17	1.05	1.01	.22
57	1.41	.15	1.09	1.12	.14
58	.92	.15	1.03	1.01	.25
59	2.27	.15	.99	.99	.29
60	1.81	.15	1.02	1.07	.24
61	-.21	.20	1.13	1.33	.02
62	.09	.18	1.05	.99	.21
63	2.83	.16	1.12	1.36	.01
64	.75	.16	1.10	1.17	.11
65	1.52	.15	1.02	.99	.27
66	-1.00	.27	.88	.84	.36
67	.05	.19	.95	1.07	.29
68	1.71	.15	1.04	1.04	.23
69	.05	.19	.95	.97	.32
70	.78	.16	.97	.99	.32
71	.75	.16	.97	.98	.32
72	-1.15	.28	.78	.39	.57
73	.97	.15	.95	.92	.37
74	-.57	.23	.91	.81	.36
75	-1.90	.39	.81	.23	.52
76	.43	.17	.96	.95	.34
77	.22	.18	.87	.80	.47
78	.85	.16	.88	.83	.48
79	.12	.18	.92	.94	.36
80	.46	.17	.97	.97	.32
81	.43	.17	.96	.97	.33
82	-.68	.24	.82	.67	.50
83	-1.15	.28	.86	.64	.42
84	-.34	.21	.83	.76	.48
85	-.09	.19	.92	.92	.36
86	1.94	.15	1.00	.97	.29
87	.28	.18	.90	.90	.41

88	-1.23	.29	.81	.56	.48
89	-1.41	.32	.78	.31	.57
90	-.74	.24	.79	.53	.56
91	-.63	.23	.79	.57	.56
92	-.86	.25	.85	.64	.45
93	.49	.17	.96	.97	.33
94	-.05	.19	.89	.89	.40
95	-.93	.26	.76	.42	.61
96	.78	.16	.97	.98	.32
97	-1.76	.37	.80	.25	.54
98	-.13	.20	.87	.81	.44
99	-1.07	.27	.75	.38	.62
100	.19	.18	.83	.81	.51