| | |
|---|---|
| العنوان: | The Effects of Test Anxiety and Item Format on Differential Item Function and Test Information Function with First Grade Secondary School Students |
| المصدر: | المجلة المصرية للدراسات النفسية |
| الناشر: | الجمعية المصرية للدراسات النفسية |
| المؤلف الرئيسي: | خضر، عادل سعد يوسف |
| المجلد/العدد: | مج23, ع80 |
| محكمة: | نعم |
| التاريخ الميلادي: | 2013 |
| الشهر: | يوليه |
| الصفحات: | 1 - 25 |
| رقم MD: | 1012056 |
| نوع المحتوى: | بحوث ومقالات |
| اللغة: | English |
| قواعد المعلومات: | EduSearch |
| مواضيع: | الوظيفية الفارقة، المواد الدراسية، طلبة الصف الأول الثانوي |
| رابط: | http://search.mandumah.com/Record/1012056 |

# The Effects of Test Anxiety and Item Format on Differential Item Function and Test Information Function with First Grade Secondary School Students

**By Dr. Adel S Y Khedr**
**Associate Prof. of Educational Psychology**
**Faculty of Education- Zagazig University - Egypt**

Detecting Differential item function (DIF) and test information function helps test constructors to avoid bias in test construction and students assessment. The aim of this study was to identify the effects of test anxiety and item format on differential item function in mathematics test and test information function. The participants were 540 first grade students from three high schools, they completed test anxiety scale and answered 40 items mathematics test: 20 items in multiple choice format (MCQ) and 20 items in true/ false format (T/FQ) prepared by the researcher. The participants were divided in two groups: low level of anxiety and high level of anxiety according to their scores on test anxiety scale using the median. Two parameters model by Biolog for M-G3 was used for the analysis. The method of detecting DIF was the differences between items' difficulties in the case of the high and the low anxiety students. In addition, the researcher employed Z test to identify the significance of percentages of DIF items between the two anxiety groups. Test information was computed as the sum of the information for the all test items (slope for each item). The findings indicated that the three items (1, 2, 19) in MCQ format showed DIF, two of them (item1 and item2) were in favor of low anxiety students and item 19 was in favor of high anxiety students. Whereas, one item (item 10) in true/false format was in favor of low anxiety students. Furthermore, using "t" test, the findings indicated that test information function varied according to item format. In addition, MCQ subtest showed more information with low anxiety students than with high anxiety students. The findings were discussed and interpreted in light of literature.

***Keywords: Test anxiety, Question format, Differential item function, Test information function.*** *Contact information: Dr. Adel S.Y. Khedr, Faculty of Education, Zagazig University, Egypt, Email.* Adelk22@yahoo.com.

# The Effects of Test Anxiety and Item Format on Differential Item Function and Test Information Function with First Grade Secondary School Students

**By Dr. Adel S Y Khedr**
**Associate Prof. of Educational Psychology**
**Faculty of Education- Zagazig University - Egypt**

## Introduction

Detecting differential item function (DIF) helps test constructors to avoid bias in test construction and students' assessment. In addition test anxiety and item type could be a source of bias in students' assessment. Furthermore, DIF items and item format may affect the amount of information interpreted by each test item and the test as whole (Baker, 2001).

Previous research showed that there are many factors that considered as a source of DIF such as test anxiety, item type (Zumbo and Gelin, 2005; Otto, et al., 2007; Kim, Naomi, and George, 2012), Familiarity with items, item characteristics, social and psychological characteristics (Nekane, et al., 2010), sample size (Pedrajita, 2009), test length, the model of the analysis (Randall & Oriel, 2010; and Khalid, 2011, ), class level and student gender (Ismail,2012).

Measuring change of psychological characteristics face many problems such as floor and ceiling effects which include using different and more difficult test to assess performance after an intervention programs, available tests not giving way of equal-interval scales, worry of using single-variable assessments and finally low level of reliability, construct validity, predicative validity and meaningfulness of differences between examinee scores before and after some treatments (Prieler and Raven 2009).

Debra et al. (2013) in his study of the international reading literacy test, indicated that item differential function could result from item cognitive loading. It was in favor of British and the American students when compared to Colombian students. In the same line of research Del Carmen (2013) found that "the Child Perceptions Questionnaire 11-14 Spanish version showed moderate-to-large uniform differential item functioning"

Using DIF approaches could help to overcome some of these

problems, help test constructors to be more accurate, faire and objective with different groups of examinees.

Despite three decades of interest in DIF, no clear picture has emerged regarding the effects of test anxiety, and item format on item information and test information functions as well as detecting DIF, particularly in Arab countries.

Pervious studies indicated a need of further study to identify the role of test anxiety and question format in detecting test information function, and differential item function.

The aims of this research are to identify the effects of test anxiety and item format on both differential item function and test information function in mathematics test.

**Research questions**

The problem of this research can be identified in the following questions.

1- Are there any items showing differential item function (DIF) according to test anxiety in mathematics test?.

2- Are there significant effects of item format on differential item function in mathematics test?.

3- Are there significant effects of item format on mathematics test information function?.

4- Are there significant effects of test anxiety on mathematics test information function?.

**Differential Item Function (DIF)**

DIF is a common way to identify test validity and to achieve fairness of assessment and testing for different groups of test takers. The basis of DIF is that two groups of test takers (a focal and a reference group) who have the same level of ability should have the same probability of answering test item correctly, DIF is present when this probability is significantly different (Kim, et al. 2010,p.326).

Furthermore, many researchers agreed that differential item functioning is a set of statistical approaches that provides signs of items that are functioning differently for different groups of students. An item shows DIF if examines of different groups having the same ability, but do not have the same probability of answering the item right (Perdrajita, 2009, Abedalaziz, 2011).

The importance of detecting differential item function and controlling these items which behave in different ways with different groups of examinees is to give equal opportunities and same advantages of all examinee groups regardless of their gender, culture, and social-economic background, psychological, environmental and behavioral characteristics.

All examine groups should have an equal opportunity to learn, to show their knowledge and to achieve their goals. Therefore, we can achieve the values of fairness, objectivity and accuracy of assessment and evaluation.

Researchers used many methods of detecting item bias such as (a) differences in relative item difficulty across different groups, (b) differences in item discrimination across groups, (c) differences in the item-characteristic curves for different groups, (d) differences in the distribution of incorrect responses for various groups, (e) differences in multivariate factor structures across groups. (Abedalaziz, 2010, p.105)

Furthermore, Tülin (2008) examined the number of DIF items detected by Hierarchical Linear Modeling at different sample sizes. The samples of the study encompassed 10727 students of 798307 are chosen by random. They were in different sample size: 97, 207, 532, 1055, 2681, 5320, 8037, 10727. The participants completed 25 items of Turkish, science, and social studies sub tests, the researcher found that varieties in sample sizes have a significant effect on DIF detection in test items.Nageep (2008), found insignificant differences between the three models of the analysis (1PL, 2PL and 3PL models) in detecting DIF. Moreover, there is a shortage of research that focused on investigating the relationship among test anxiety, item type, the model of the analysis and test information function.

In the same line of research, Pedrajita (2009) used logistic regression analysis to detect differential item function between public and private school students as well as males and females students at these schools, with 200 students, 100 students from each school, 100 students from each gender group. The participants answered a chemistry achievement test. The findings indicated that 11 items showed bias in favor of public school, and 11 items biased in favor of private school. 4 items showed bias against male students and 3 items showed bias against females' students. These findings raise some questions related to sample size used.

Randall and Oriel (2010) investigated the differential item functioning in its relation to ethnical groups and test length. They used the standard 20-items scale, 10-items and 12-items subsets of the Center for

Epidemiological Studies Depression Inventory. With hispanics and the non-hispanic whites individuals. They also compare item use among the elderly Hispanics over time. The participants were 4499 elderly individuals. Partial correlation analysis was applied to test hypotheses regarding the DIF on the basis of ethnic background and among the elderly Hispanics across time points. The findings indicated that Hispanics appear to express positive affect differently than non-Hispanic Whites. On the 20-items scale, Hispanics under responded to items measuring positive affect. Item bias was also found for one positive affect question in the 10-item subset. Among the Hispanic sub-population, it is also noted that item response bias for the positive affect items across time. A 12-items subset tested for DIF displays an acceptable amount of item bias across ethnic groups and across time for the Hispanic sample.

More recently Abedalaziz (2011) used item characteristic curve approach to detect DIF items in mathematics test related to gender, Model of the analysis (1PL, 2PL models) and content areas within mathematics. The participants included 1318 tenth grade students (664 males and 654 females). They completed achievement test of tenth grade covering the subjects of relations and functions, polynomial, trigonometric functions, and triangles. Results indicated that: the differences between the two models in detecting DIF is relatively high. In addition to that there is a gender- related DIF linked to content areas within mathematics. The nature of mathematics context related to detecting DIF items.

Taylor and Lee (2011) investigated ethnic differential item function for 4th $7^{th}$ and $10^{th}$- grade in reading test. The test items were in multiple choice and constructed response items formats. The participants were in two different groups: White and non-Wight students. The findings indicated that multiple-choice items were in favor of white students, whereas constructed response items were in favor of non- white students. Furthermore, DIF items in both formats measure literal comprehension as an essay reading skill, as well as developing interpretations across text and analyzing graphic elements as higher level skills of reading. Kim, Naomi and George (2012) investigated differential item function (DIF) of test items on Pennsylvania's Altemate System of Assessment with visual impairments and severe cognitive disabilities. The findings indicated that among the functional vision groups, 17 items in reading and 22 items in math showed DIF. Furthermore, 14 skills in reading and 13 in math were harder for students with visual impairments. Finally DIF items varied according to item format.

Ogundokun (2011) investigated learning styles, school environment and test anxiety as predictors of learning outcomes with three hundred secondary school students. The participants were administered four valid and reliable instruments: learning styles, school environment and test anxiety. Pearson's Product Moment Correlation and Multiple Regression Analysis were used to analyze the data. The findings indicated that learning styles, school environment and test anxiety together predicts the learning outcomes but test anxiety is the strongest predictor of learning outcomes.

According to previous research, it was noticed that DIF affected by many factors such as subject content, Item formats, sample size, model of the analysis, test length, and test takers psychological and environmental factors. The focus of this research is on some of these factors such as, item formats, level of students' test anxiety.

## Test Anxiety

Test anxiety is a common problem examinee face. It may have a negative effect on their academic achievement; therefore, it is important to identify this problem and their effect on test taker test's performance. One fundamental problem faced test takers in secondary education is test anxiety, this anxiety is appear as a set of negative emotional signs and a high degree of worry related to their expectations of unsatisfied performance on tests and examinations.(Albert, Alfonso, and Javier, 2010). Perry, (2004), defined test anxiety as "an otherwise confident student's state of panic during a test where self-doubt leads to a failure to realize potential in a testing environment" (p. 321).

New trends in assessment system aimed to identify likely test anxiety problems in examinees and will permit the enhancement of their test performance. Some researchers studied test anxiety as relatively stable trends in individuals to produce anxiety responses when facing test situations, separate from a measurement of general anxiety.(Albert, Alfonso, and Javier, 2010).

Wilder (2012) distinguish between two types of test anxiety: "trait anxiety" which referred to as an enduring and continual tendency to be anxious. It is in common related to individual who have an anxiety disorder. These individual could become more anxious and exhibit some negative

behavior in some situations. Whereas, "State anxiety" is defined as an enduring tendency to be anxious in particular situations, the individual shows this type of anxiety when facing scary stimulus or frightening situations. The source of test anxiety could be poor preparation; this includes lack of time management skills, over study the night before the exam, poor study habits, second, it could be immerge as a worrying about previous test answers; how students schoolmates doing in the same test and the negative consequence of their results. Furthermore, some students might exhibit bodily symbols of test anxiety during the exam situation.

Some researchers studied test anxiety as a one dimensional phenomenon, whereas others studied test anxiety as a multidimensional one. For example Hodapp (1991); Heredia, Piemontesi, Furlan, and Hodapp (2008); and Albert, Alfonson and Javier, (2010), investigated test anxiety as a multycomponent construct. It includes emotionality, worry, interference and lack of confidence. If test anxiety left without intervention, it could result in negative consequences. Such as poor test performance, negative self-steam and negative self concept as well as low and poor level of test information related to the content of students abilities (Knox et al. 1993).

Test anxiety is the strongest predictor of test performance, in comparison with learning style, school environment (Ogundokun, 2011). It has a negative effect on mathematics test performance (Engelhard, 2001,). Furthermore, some researchers found negative correlation between some dimensions of anxiety (lack of confidence and interference) and students academic performance (Heredia et al. 2008 and Albert et al. 2010). On the other hand, Birjandi and Alemi (2010) found mixed relationships (positive and negative correlations) among test anxiety and test performance. But the researchers concluded that "test anxiety does not seem to much influence on test performance at the micro- test-specific level".

Putwain (2008) with British students found a significant negative correlation between test performance on GCSE examinations and test anxiety. Furthermore, cognitive component of test anxiety accounts for 7% of variance in examination performance. Recently, Wilder (2012) indicated to the role of gender differences between correlations of age, math achievement, perceived, and teacher attitude towards mathematics and math anxiety.

Many types of test anxiety affects on individual such as nervousness, mental blocking and common worries (Wilder 2012) these factors affect the amount of information interpreted for each item and the test as a whole.

Therefore, identifying the role of test anxiety in differential item function could result in achieving test fairness. Test constructers should insure that test items free of bias agents or in favor of low or high anxiety students.

Previous studies ignored the role of test anxiety in relation to item differential function. This is one focus of this study. Furthermore, Pervious results indicated to a need of further study to identify the role of test anxiety in detecting item information function, differential item function and the amount of information interpreted by the test,

## Test Information Function.

Item and test information function are important and useful in constructing tests and examinations and achieves test fairness and accuracy (Hambelton and Jones,1994). The term information defined by Fisher (1968) as the mutual of accuracy with which item parameter could be estimated.

Information shows that how every ability level has been calculated. The information function shouldn't be recognized to be relied on the distribution of examinees over the ability scale. The ideal information function would be horizontal line at the same large value of information (I) and all ability levels would be calculated precisely. The amount of item information dwindles as the ability level will be different from the item difficulty and comes near zero at the extremes of the ability scale. (Baker, 2001, p.105-107).

The researcher used the total information area index to indicate the area under the total information function, since the items contribute independently to the total information function of the test. The area under the total information function is the sum of the item information areas (de Ayala, 2009, p.115). In the case of one (1PL). Test information function can be computed as:

$$I_A = N\alpha \quad \text{where,} \qquad (1)$$

$I_A$= the total information of the test. N= number of items, and $\alpha$= item information (Slope for each item).

Whereas In the case of two parameters model (2PL). The total information of the test can be computed by the equation:

$$I_A = \sum \propto_i \qquad (2)$$

Where $I_A$= the area under the total information function,

$\sum \propto_j$ = the sum of the information for the all test items j (Slope for each item).

The overall test information function equal to the simple summation of item information of all test items: $I(\theta)= \sum_{k=1}^{n} I_j(\theta)$ (3)

(Baker, 2001, p. 239-243)

Standard error of test information is referred to by the following equation

$$SE\ (I(\theta))=\frac{1}{\sqrt{I(\theta)}} \quad (4)$$

The standard error of test information refers to the estimation accuracy of test information function. (Hambleton & Swaminathan, 1985)

The amount of information depends upon the particular item parameters (Baker, 2001). Some arguments suggested that sample size and test length influence on the accuracy of item and test information functions. Furthermore, errors in item parameter estimates affect negatively on the accuracy of item and test information functions.

Test information function relayed on the number of items in the test, the average values of discrimination parameters of the test items. Furthermore the shape of the test function depends upon the distribution of item difficulties over the ability scale, as well as the distribution and the average value of the discrimination parameters of the test items. Baker (2001, P.127).

Previous studies ignored the effect of test anxiety and item format on differential item function and test information function. Therefore, there is a need to conducting more investigation in order to shed more light on this issue.

## Hypotheses

1-There are some items showing differential item function (DIF) according to test anxiety in mathematics test.

2- There are significant effects of item format on differential item function in mathematics test.

3- There are significant effects of item format on mathematics test information function.

4- There are significant effects of test anxiety on mathematics test information function.

### Methodology

#### Sample

The Pilot sample encompasses 106 (96 valid cases) of first grade high school students; they completed a test anxiety scale and answered mathematics test. Their ages ranged from (15) to (18) years with mean age = 15.78 and SD= 0.66. The final sample included (540) first grade students from three high schools, their ages ranged from 14 to 21 years with mean age = 15.85 and standard deviation (SD=0.89).

## Instruments

### 1- Test Anxiety Scale

### Reliability

The researcher prepared a test anxiety Scale which encompasses (27) items for Arabic culture by reviewing previous scales and researches. This scale was reviewed by some educational psychology specialists. The primary version of the scale was completed by 106 high school students. Item reliability was computed in the case of excluding each item from the total score. Item reliability extended from 0.867 to 0.885. Scale reliability was 0.87, and the findings indicated 3 unreliable items. They were excluded and Reliability Alpha after excluding unreliable items was = 0.89. The relation between test anxiety scale and achievement in mathematics = 0.014

***Validity:*** explanatory factor analyses (Extraction Method: Principal Component Analysis and Rotation Method: Varimax with Kaiser Normalization) was performed on the data. The findings indicated five

factors which interpreted 56% of variance. The mean roots for these factors are: 2.24, 1.96, 1.90, 1.8, and 1.08 respectively. Confirmatory factor analysis was performed on the correlation matrix. The findings indicated that one factor model solutions indicated a goodness of fit (chi-square= 4.59, df=5, P=0.47, RMSEA= 0.000). The loadings of the five factors on the latent factor were 0.76, 0.77, 0.59, 0.73, 0.49.

## 2- Mathematics Criteria Referenced Test

The Mathematics criteria referenced test encompasses (40 Items) - 20 multiple choice items, and 20 True - False items in the unit of sequences prepared by the researcher using criteria referenced test procedures.

**Test Reliability**: reliability Alpha for the main test, multiple choice question sub test and true/false questions subtest were 0.81, 0.86, 0.72 respectively. Kudar&Richardson reliability coefficient were computed. These coefficients were 0.86, 0.90, and 0.70 respectively.

## *Test Validity.*

The researcher used criterion validity by computing the correlations between the main mathematics test, and sub tests with academic achievement in mathematics. These correlation coefficients were 0.89, 0.94, and 0.48; respectively.

## Internal consistency

The researcher computed the correlation between Multiple choice questions and the total score (r= 0.92) and between true / false questions and total score (r= 0.74), and between multiple choice and true/false questions (r= 0.433). These findings refer to the reliability, validity and internal consistency of the instruments.

## Procedures:

1- The researcher conducted literature review
2- Test anxiety scale was developed and standardized with a pilot sample
3- The Mathematics Test was constructed and standardized with a pilot sample, using SPSS and Liseral 8.8.
4- The participants completed test anxiety scale and answered The Mathematics Test in separate sessions. Mathematics test consists of 40 items, 20 items in multiple choice question (MCQ) and 20 items in true /false question (TFQ).

5- The Test Anxiety Scale and the Mathematics Test was completed and scored by the researcher.
6- The participants were in two groups: low level of anxiety and high level of anxiety according to their scores on test anxiety scale.
7- The researcher used Bilog MG-3 program to analyse the data.
8- The hypotheses were tested and the findings were discussed in the light of literature.
9- Conclusion and recommendations were presented.

## Data Analysis and Findings.

## First hypothesis

To test the first hypothesis which states that "There are some items showing differential item function (DIF) according to test anxiety in Mathematics Test", the researcher used 2LP model by employing Biolog-MG3.

The method of detecting DIF was the differences between items' difficulties in the case of the high and the low anxiety students. (Camilli & Shepard,1994), when $(\theta)$ pramter is constant. The differential item function computed by equation number (5)

$$\Delta_b = (G2 - G1)/\text{SE} \qquad (5).$$

$G2$ = *Item difficulty for study group*

$G1$ = *Item dificulty for focal group*

$\Delta_b$ = *The differences between items difficulty for two groups*

SE= Standard Errors

if $\Delta_b > 1.96$ then item shows DIF at 0.05 and If $\Delta_b > 2.58$ then item shows DIF at 0.01. Table (1) shows these findings for multiple choice questions (MCQ) when using 2PL.

Table(1). The differences between items' difficulty for low and high anxiety students as an indication of DIF for MCQ using 2PLModel

| Items | G2-G1 Thresholds | S.E | (G2-G1)/S.E | Item DIF | Items | G2-G1 | S.E | (G2-G1)/ S.E | Items DIF |
|---|---|---|---|---|---|---|---|---|---|
| Item1 | -0.206 | 0.097 | -2.124 | DIF | Item11 | -0.098 | 0.242 | -0.405 | NDIF |
| Item2 | -0.317 | 0.133 | -2.383 | DIF | Item12 | 0.164 | 0.226 | 0.726 | NDIF |
| Item3 | -0.076 | 0.142 | -0.535 | NDIF | Item13 | 0.002 | 0.132 | 0.015 | NDIF |
| Item4 | -0.067 | 0.071 | -0.944 | NDIF | Item14 | -0.23 | 0.143 | -1.61 | NDIF |
| Item5 | -0.003 | 0.068 | -0.044 | NDIF | Item15 | -0.091 | 0.176 | -0.517 | NDIF |
| Item6 | -0.219 | 0.13 | -1.685 | NDIF | Item16 | 2.054 | 1.839 | 1.117 | NDIF |
| Item7 | -0.289 | 0.179 | -1.615 | NDIF | Item17 | -0.437 | 0.321 | -1.361 | NDIF |
| Item8 | -0.057 | 0.111 | -0.514 | NDIF | Item18 | -0.136 | 0.124 | -1.097 | NDIF |
| Item9 | -0.086 | 0.132 | -0.652 | NDIF | Item19 | 0.48 | 0.232 | 2.069 | DIF |
| Item10 | -0.189 | 0.143 | -1.322 | NDIF | Item20 | -0.198 | 0.249 | -0.795 | NDIF |

Table (1) shows that item number (1) and item number (2) showed DIF in favor of low anxiety students. Whereas, item number (19) showed DIF in Favor of high anxiety students.

Table (2). The differences between items' difficulty for low and high anxiety students as an indication of DIF for T/FQ using 2PL Model.

| Items | G2-G1 Thresholds | S.E | (G2-G1)/S.E | Item DIF | Items | G2-G1 | S.E | (G2-G1)/S.E | Items DIF |
|---|---|---|---|---|---|---|---|---|---|
| Item1 | -0.703 | 0.981 | -0.717 | NDIF | Item11 | -0.061 | 0.224 | -0.272 | NDIF |
| Item2 | 0.17 | 0.271 | 0.627 | NDIF | Item12 | -0.281 | 0.986 | -0.285 | NDIF |
| Item3 | 0.11 | 0.389 | 0.283 | NDIF | Item13 | 0.326 | 1.076 | 0.303 | NDIF |
| Item4 | 0.95 | 0.894 | 1.063 | NDIF | Item14 | -0.212 | 1.218 | -0.174 | NDIF |
| Item5 | 0.393 | 0.353 | 1.113 | NDIF | Item15 | -0.295 | 1.152 | -0.256 | NDIF |
| Item6 | 0.034 | 0.202 | 0.168 | NDIF | Item16 | -0.299 | 0.619 | -0.483 | NDIF |
| Item7 | -0.354 | 0.717 | -0.494 | NDIF | Item17 | 0.214 | 0.468 | 0.457 | NDIF |
| Item8 | 0.337 | 0.44 | 0.766 | NDIF | Item18 | -0.262 | 0.418 | -0.627 | NDIF |
| Item9 | 0.123 | 0.358 | 0.344 | NDIF | Item19 | 0.483 | 1.472 | 0.271 | NDIF |
| Item10 | -0.328 | 0.151 | -2.172 | DIF | Item20 | -0.345 | 0.860 | -0.401 | NDIF |

Table (2) shows that item number (10) of true false questions (T/FQ) is in favor of low anxiety students. The findings indicated that 3 items in Multiple choice format show DIF. Items (1,2) were in favor of high anxiety students, Whereas item 19 were in favor of low anxiety students, and one item in true/false format (item number 30) was in favor of high anxiety students.

## Second Hypothesis

To test the second hypothesis which state that, "there are significant effects of item format on differential item function in mathematics questions test"

The researcher employed "Z" test to identify the significant between items formats in differential item function in mathematics test according to item formats. Critical Values for Z extend from 1.96 to 2,58. Table (3) shows the number of DIF items by question formats.

Table (3). Number and percentage of DIF items by item format and model of the analysis.

| | Number of DIF items. Two Parameters Model | Percentage |
|---|---|---|
| M-C - Items | 3 | 15% |
| T-F- Items | 1 | 5% |
| Z | | 0.26 |

Z test of percentages significant showed that, there is insignificant difference between the percentages of DIF items according to item formats multiple choice questions (MCQ) and true false questions T/FQ).

## Third Hypothesis

To test the third hypothesis which state that "There are significant effects of item format on test information function", the researcher used t-Test to identify the significance of the differences between means of information interpreted by each subtest. The means and SD for each subtest were computed using Bilog MG3.

The findings indicated that the differences between the mean of information interpreted by multiple choice questions and true false quests were significant at 0.01. Table (4) shows these findings.

Table (4). Mean, SD and "T" test for the information of MCQ question and T/FQ

| | M | SD | N | t- test vale |
|---|---|---|---|---|
| MCQ subtest | 1.866 | 1.413 | 20 | 3.19** |
| T/FQ subtest | 0.874 | 0.485 | 20 | |

** Significant at 0.01.

In addition, the researcher used the method suggested by (de Ayala, R.J., 2009,p.116-117) to compute the total information interpreted by multiple choice question( MCQ) subtest and true/false question (T/FQ) subtest. The researcher used equations number (1) and number (3) to calculation the amount of information for each test.

This approaches relied on adding up the items information in each test to get the total information for that test. Table (5) shows these findings.

Table (5). Item format and test information function.

| Two P. Model | MCQ subtest | T/FQ subtest | MCQ Information/T-F Q information (de Ayala, R.J., 2009,p.116-117) |
|---|---|---|---|
| **Total Information** | 37.315 | 17.475 | 37.315/17.475= 2.14 |

Table (5) indicated that the amount of information interpreted by multiple choice questions subtest is 50% more that that information interpreted by true/false questions subtest.

## Fourth Hypothesis

To test the fourth hypothesis which state that, "There are significant effects of test anxiety on mathematics test information function", the researcher used the plots for each test formats for high and low anxiety group. Tables 6, 7 and figures 1, 2,3,4 show these finding.

The following figures show test information plots by item format, for low anxiety students using 2PL model.

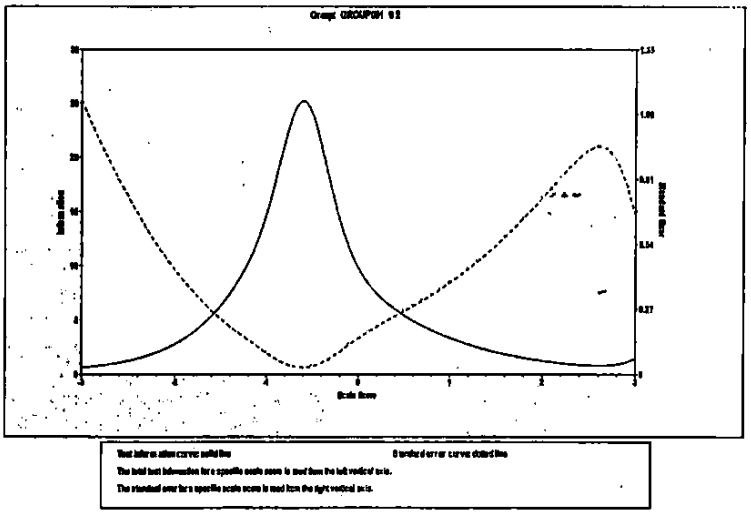

Figure (1) Information for low anxiety students in the case of MCQ

Figure (1) and table (6) show that in the case of low anxiety students the highest information of the (MCQ) questions (subtest) was at the interval (0.4, -1.4). The test interprets 26% of information at $(\theta)$ = -0.6. This was at the lowest level of standard error at this value of $(\theta)$.

Table (6). Test anxiety and highest level of Information for MCQ

| **Multiple choice question** | | | |
|---|---|---|---|
| Anxiety Level | Interval of information | Highest level of information | $(\theta)$ |
| High anxiety | From 0.3 to -1.8 | 24 | -.7 |
| Low anxiety | From 0.4 to -1.4 | 26 | -.7 |

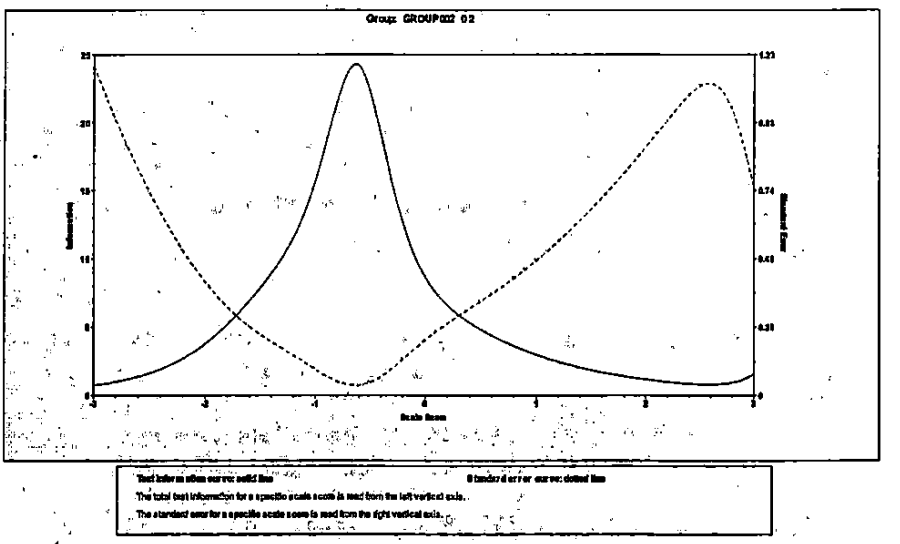


Figure (2) Information for High anxiety students in the case of MCQ

Figure (2) and table (6) show that in the case of high anxiety students, the highest information of multiple choice questions (MCQ) (subtest) was at the interval (0.3, -1.8). The test interprets 24% of information at $(\theta)$ = -0.7. This was at the lowest level of standard error at this interval.

Figure (1), figure (2) and table 6 indicated that multiple choice questions subtest showed more information with low anxiety students than with high anxiety students.

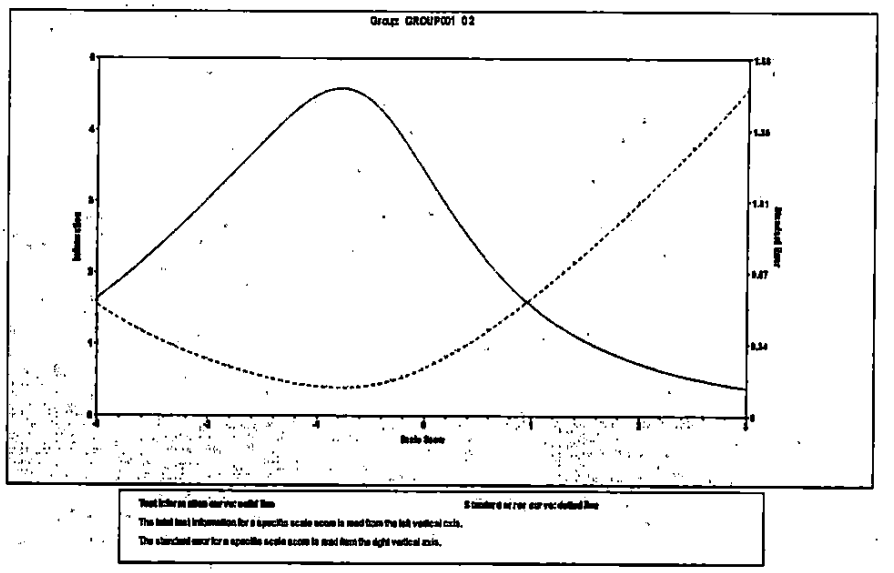


Figure (3) Information for low anxiety students in the case of T/FQ

Figure (3) and table (7) show that in the case of low anxiety students the highest information of the true/ false questions (T/FQ), was at the interval (-0.5, -1.5). The test interprets 4,6% of information at $(\theta) = -1.2$ . This was at the lowest level of standard error at this interval. True/false questions showed higher information in the case of higher anxiety students.

Table (7). Test anxiety and the highest level of Information.

| True False question | | | |
|---|---|---|---|
| Anxiety Level | Interval of information | Highest level of information | $(\theta)$ |
| High anxiety | From -0.1 to -1.8 | 4.6 | -1.4 |
| Low anxiety | From -0.5 to -1.5 | 4.6 | -0.8 |

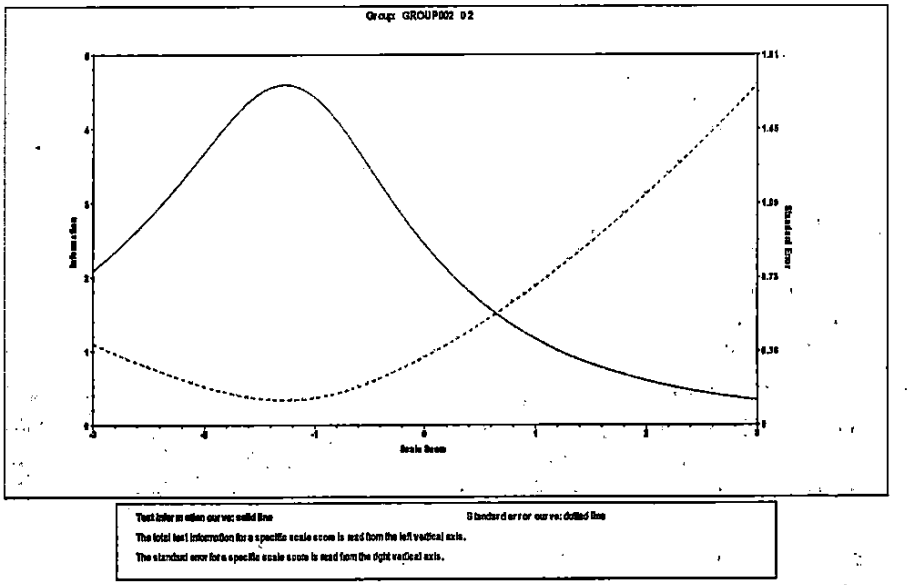


Figure (4) Information for low anxiety students in the case of T/FQ

Figure (4) and table (7) show that in the case of high anxiety students the highest information of the true/ false questions (T/FQ) was at the interval (-1, -1.8). The test interprets 4,6 % of information at $(\theta)$=-1.4 This was at the lowest level of standard error at this interval.

Figure (3), figure (4) and table (7) show that test information function of true false questions is the same with both low and high anxiety students. Therefore, there is no difference between low and high anxiety students in test information function, when using True/false question subtest.

## Discussion and Conclusion

The findings indicated that some items showed Diff according to test anxiety and item formats. Two items in multiple choice formats (Item number 1 and item number 2) showed DIF in favor of high anxiety students, whereas one Item (item19) showed DIF in favor of low anxiety students. In addition, one item in true false format showed DIF in favor of high anxiety students.

Although these findings indicated that there is a variety of items that showed differential item function according to test anxiety and test formats,

No significant differences between the percentages of items showing DIF according to test question format have been found. These findings are consistent with the findings of Randall and Oriel (2012), who found that -in a short test of 12 items- some items showed an acceptable amount of items bias across ethnic groups. Also our findings are consistent with the findings of Debra et al. (2013), they found that item cognitive loading in the international reading literacy test was in favor of British and the American students when compared to Colombian students. Whereas, the findings of current research are inconsistent with the findings of Taylor and Lee (2011) and Kim, et al. (2012).They found that differential item function affected by question format. May be the inconsistency is due to the differences between the subjects used and skills being measured in their studies.

The findings indicated that there are significant differences between multiple choice and true/ false questions in test information function. These findings support the idea that multiple choice questions give more information than true false questions do. This finding support the findings of Nekane, et al.(2010) who indicated that item characteristics, social and psychological characteristics affects on students test performance. Furthermore, Taylor and Lee (2011) revealed that multiple-choice items were in favor of white students, whereas constructed response items were in favor of non- white students. Moreover, DIF items in both formats measure literal comprehension as an essay reading skill, as well as developing interpretations across texts and analyzing graphic elements as higher level skills of reading.

The findings of this study referred to small number of test items showed Dif (4) items in both formats. Whereas, Kim, Naomi and George (2012), in their study found that 17 items in reading and 22 items in math showed DIF. DIF Items varied according to item format.

Although, the findings showed that there is an insignificant difference between the percentages of DIF items according to item formats; test constructors should make a balance between different formats of test questions to meet different cognitive abilities of test takers.

The findings indicated that multiple choice question subtest showed more information with low anxiety students than with higher anxiety students. Whereas, true/false questions subtest give the same information with both high anxiety students and low anxiety students. This finding is consistent in part with the finding of Ogundokun (2012) that showed that test anxiety is the strongest predictor of learning outcomes.

In conclusion, item format and test anxiety could be a source of bias in tests and examinations. Furthermore, test information function varied

according to item type. More research in this area is needed particularly on a large scale of students. Many test questions formats (multiple choice, construct response questions, problem solving questions) and many models of IRT analysis with different subjects such as English, Science, and Arabic also need more investigation. Furthermore, considering other factors such as measurement accuracy indicators, item parameters and students' abilities are needed particularly in Arab countries.

# References


Abedalaziz, A .(2010). A Gender- Related Differential Item Functioning of Mathematics Test Items, *The International Journal of Educational and Psychological Assessment, Vol. 5, 101-116.*

Abedalaziz, N. (2011). Detecting DIF using Item Characteristic Curve Approaches, *The International Journal of Educational and Psychological Assessment, Vol. 8(2)*

Albert S.; Alfonso, P. and Javier. P. (2010). Construct validation for the German test anxiety inventory Argentinean version (GTAI-A). in a Spanish population. Cognition, Brain, and Behavior, ISSN: 1224-8398, Volume XIV, No. 4 (December), 413-429.

Baker, P.B. (2001). The Basics of item response theory,(2th ed), available from http://ericae.net/irt/baker.

Birjandi, P. and Alemi, M. (2010). The Impact of Test Anxiety on Test Performance among Iranian EFL Learners, *BRAIN. Broad Research in Artificial Intelligence and Neuroscience Volume 1, Issue 4,*ISSN 2067-3957.

Camilli, G. & Shepard, L. A. (1994). Methods for identifying biased test items. California: Sage Publications.

De Ayala, R. J.(2009). The theory and practice of item response theory, The Guilford Press, New York, London.

Debra, S; Maria, E.; Bruno, D. and Kadriye, E. (2013). Investigating Sources of Differential Item Functioning in International Large-Scale

Assessments Using a Confirmatory Approach, *Journal of international testing.* *Vol.13,* *Issue2,* *152-174*

Del Carmen, A. F.; (2013). Differential item functioning of the Spanish version of the Child Perceptions Questionnaire. *J Investig Clin Dent.* *Feb;4(1):34-8*

Engelhard, G. (2001). Math anxiety, mother's education, and the mathematics performance of adolescent boys and girls: Evidence from the United States and Thailand. *The Journal of Psychology, 124*(3), 289-298.

Fischer, G. (1968). Psychologische test theorie [Psychological test theory]. Bern: Huber.

Hambleton, R.K. and Swaminathan, H.(1984). Item response theory. Principles and applications. Hingham, MA; Kluwer, Nijhoff.

Hambleton, R. K.; and Jones, R. W. (1994). Item Parameter Estimation Errors and Their Influence on Test Information Functions, *Applied Measurement in Education, 7(3), 171-186.*

Heredia, D., Piemontesi, S., Burlan, L., & Hodapp, V. (2008). Adaptación del Inventario Alemán de Ansiedad ante los Exámenes: GTAI-A [German Test Anxiety Inventory Adaptation: GTAI-A*]. Evaluar, 8, p. 46-60.*

Hodapp, V. (1991). Das Prüfungsängstlichkeitsinventar TAI-G: Eine erweiterte und modifizierte Version mit vier Komponenten [The Test Anxiety Inventory TAI-G: An expanded and modified version with four components]. *Zeitschrift für Pädagogische Psychologie,5*,121-130

Ismail K. (2012). An Investigation of Item Bias in Science & Technology Subtests and Mathematic Subtests in Level Determination Exam (LDE). Educational Sciences: Theory & Practice - 12(1). Winter.

222-229, Educational Consultancy and Research Center.

Kim T. Z., Naomi Z., and George J. Z. ( 2012 ). Ability or Access-ability: Differential Item Functioning of Items on Alternate Performance-Based Assessment Tests for Students with Visual Impairments. *Journal of Visual Impairment & Blindness*.

Nageep, E. (2008), Methods of Item Bias Detection: A Comparative Psychometric Study Master thesis, Ain Shams University, Cairo, Egypt.

Nekane B., Arantxa G., Juana G-Benito and María D. H.(2010). Use of multilevel logistic regression to identify the causes of differential item functioning. *Psicothema. Vol. 22, n4, pp. 1018-1025*

Ogundokun, M.O.(2011). Learning style, school environment and test anxiety as correlates of learning outcomes among secondary school students. *Ife Psychologia; 19(2). P. 1-17*

Otto B. Walter; Janine B.; Jakob B. B.; Herbert F.; Burghard F. K.; and Matthias R.(2007 ). Development and evaluation of a computer adaptive test for 'Anxiety' (Anxiety-CAT). *Qual Life Res 16:143–155.*

Pedrajita, J. Q.(2009). Using logistic regression to detect biased test items. *The International Journal of Educational and Psychological Assessment, Vol. 2, 54-73.*

Perry, A.B. (2004). Decreasing math anxiety in college students. *College Student Journal, 38*(2), 321-324.

Putwain, D., W. (2008). Test anxiety and GCSE performance: the effect of gender and socio-economic background, *Educational Psychology in Practice, Vol. 24, No. 4, 319–334*

Randall C. M. and Oriel J. S. (2010).Differential item responses on CES-D inventory: A comparison of elderly Hispanics and non-Hispanic

Whites in the United States and item usage by elderly Hispanics across time. *Aging & Mental Health, Vol. 14, No. 5, July 2010, 556–564,*

Taylor, C. S. and Lee, Y. (2011). Ethnic DIF in Reading Tests With Mixed Item Formats. *Educational Assessment, 16:35–68, 2011.*

Tülin, A. (2008). Sample Size in Differential Item Functioning: An Application of Hierarchical Linear Modeling. *Educational Sciences: Theory & Practice* - 11(1) . 284-288.

Wilder, S. (2012). Gender differences in factors pertaining to math anxiety among college students, A Dissertation Presented to The Graduate Faculty of The University of Akron. USA.

Zumbo, B.D., & Gelin, M.N. (2005). A matter of test bias in educational policy research: Bringing the context into picture by investigating sociological / community moderated (or mediated) test and item bias. *Journal of Educational Research and Policy Studies, 5. 1-23.*

# أثر كل من قلق الاختبار وصياغة المفردة على الوظيفية الفارقة للمفردة ودالة معلومات الاختبار لدى طلبة الصف الأول الثانوي

د/ عادل سعد يوسف خضر
قسم علم النفس التربوي
كلية التربية- جامعة الزقازيق

ملخص الدراسة

اكتشاف الوظيفية الفارقة للمفردة ودالة معلومات الاختبار يساعد مصممي الاختبارات لتجنب التحيز في بناء الاختبارات وتقييم الطلبة. ويهدف البحث الحالي إلى دراسة أثر قلق الاختبار وصياغة المفردة على الوظيفية الفارقة للمفردات ودالة معلومات الاختبار وذلك في اختبار مرجعي المحك في مادة الرياضيات. شملت عينة التقنين (١٠٦) من طلبة الصف الأول الثانوي، كما تكونت العينة النهائية من (٥٤٠) طالباً من طلاب الصف الأول الثانوي طبق عليهم مقياس قلق الاختبار من إعداد وتقنين الباحث، واختبار مرجعي المحك في مادة الرياضيات مكون من ٤٠ مفردة: ٢٠ مفردة اختيار من متعدد، ٢٠ مفردة صح وخطأ في وحدة المتتابعات من إعداد وتقنين الباحث. تم تقسيم العينة إلى مجموعتين مرتفعي القلق ومنخفضي القلق بناءً على استجاباتهم لمقياس قلق الاختبار باستخدام الوسيط. وبتوظيف النموذج ثنائي البارامتر، باستخدام برنامج Biolg MG3 واختبار Z واختبار "ت" توصل الباحث إلى أنه بالنسبة لأسئلة الاختيار من متعدد فقد أظهرت ثلاث مفردات وظيفية فارقة اثنين منها لصالح منخفضي القلق وواحدة لصالح مرتفعي القلق. في حين وجد أن مفردة واحدة من مفردات الصواب والخطأ متحيزة لصالح منخفضي القلق. كما أشارت النتائج إلى أنه يوجد فروق دالة بين أسئلة الاختيار من متعدد وأسئلة الصواب والخطأ في دالة معلومات الاختبار. حيث وجد أنه في حالة أسئلة الاختيار من متعدد فإن دالة معلومات الاختبار لدى منخفضي القلق أعلى منها في حالة مرتفعي القلق. في حين لا توجد فروق في دالة معلومات الاختبار بين مرتفعي القلق ومنخفضي القلق في حالة أسئلة الصواب والخطأ. وقد تم مناقشة وتفسير النتائج في ضوء أدب البحث.