European Scientific Journal December 2014 edition vol.10, No.35 ISSN: 1857 – 7881 (Print) e - ISSN 1857- 7431

# THE IMPACT OF A NUMBER OF CHOICES IN A TEST ON THE DIFFICULTY INDICATORS AND DISCRIMINATION

***Ahmed Atallah Ali Alsaraireh***
***Heba Suliman Alrawashdeh***
***Dr. Adnan Alrawashdeh***
The University of Jordan – Jordan

**Abstract**

This study aimed to investigate the relationship between a number of choices (alternatives) in a test, and the difficulty indicators and discrimination. A test of multiple choices on statistics students at the University of Jordan was used in this study. The test consisted of 30 items, and three forms (models). Also, a sample of 350 students was studied. The Statistical Package for Social Sciences (SPSS) software and Bilog-Mg were used for the analysis of the data based on the Two parameter models. Therefore, the findings of the present study uncovered a negative association between the number of choices and the difficulty indicator and discrimination.

**Keywords:** Difficulty, Discrimination, test, Two parameter models, University

## Introduction

The educational field has reflected a growing development in the design, construction, and analysis of the item. One of these theories which were used in this study is the classical theory of measurement (CCT). This theory has been used for a long time, but this theory was dependent on the sample scores. However, the relation between the properties of the test and the students are not independent (Hambleton and Swamnthan (1985 )). Furthermore, the educational field made use of the new theory. This theory is called Item Response Theory (IRT), and it assumes the existence of common characteristics among the students who have been tested.

The new theory has a mathematical model to estimate the properties of items and test takers. Thus, these properties are invisible, and so, it is called Latent Trait (Crocker & Algina, 1986).

This field has increasing developments in the design, construction, and analysis of test items to achieve this objective. The tests were considered the most important tools for data collection and decision-making; for instance, choosing someone for a job, or the evaluation of student in classroom tests **(Allen & Yen, 1979).**

The choice tests are the most popular in the world because they are efficient, flexible, and easy to check; thus we can measure the targets in various subjects of this study **(Aikan, 1987)**. The test was divided into two parts. The first part is the question, while the second part consists of alternatives, which contains only one correct answer. The alternatives are homogeneous **(Nitko, 2001),** and the properties that control the performance and ability of the students include: the homogeneity of the alternatives, the difficulty of the question, guessing factor, and the clarity of the question **(Plake, Thompson & Lowry, 1981).**

The Previous studies have shown a lack of agreement on a certain number of choices. The study by Costin (1970**)** shows that the average discrimination coefficients and the average of difficulty coefficient of the test items with three choices (alternatives) are higher than the average discrimination coefficients and average of difficulty coefficient of the test items with four choices (alternatives). Straton **&** Cats, (1985) stated that the number of alternatives is inversely proportional to the items difficulty, and the item discrimination is directly proportional to the number of choices. Also, Crehan **&** Haladyna, (1993) stated that the form with three choices saves time and effort, and the average of difficulty coefficient of the three choices is greater than the average of difficulty coefficient form with four choices.

**Rodriguez (2005)** analyzed studies which were conducted through (80) years, and the results showed that the test of the three choices was the best. It also shows that the change of the five alternatives to the four alternatives works to decrease the difficulty of item to about (0.02). The coefficient of discrimination was about (0.04), and the coefficient of stability was about (0.035).

In **Shizuka, Takeuchi, Yashima, Yoshizawa, (2006)**, the results showed that the coefficients of difficulty and discrimination with respect to all forms were not statistically significant. Thus, it has already shown that there were variations in the results of the studies. However, this study was conducted to finish what has begun.

**The Problem of the Study and Questions**

Due to the fact that we do not arrive at a final decision from the previous studies about the best number of choices, we conducted this study at the University of Jordan on students studying statistics. Also, we depended

on the new theory to analysis the results of the response to the three models. The new theory of items estimates is free from the characteristics of item properties. Therefore, through this study, we will try to answer the following questions:

**Question 1**: Is the accuracy estimate of the difficulty indicator by two parameters model depending on the number of choices (alternatives)?

**Question 2**: Is the accuracy estimate of the discrimination indicator by two parameters model depending on the number of choices (alternatives)?

**Importance of Study**

This study is aimed to investigate the effect of the number of choices (alternatives) on the accuracy of estimating the difficulty indicator and discrimination of the test items and information function for the test. Thus, we estimate the properties of items (difficulty and discrimination) in independent way with respect to the measure of student's achievement in statistics at the University of Jordan.

**Definition of Terms**

**Multiple Choice Test:** Every item has two parts: the first part is called the question, while the second part is called the choices (alternative). The examiner takes two marks i.e. Zero (0) if the answer is false, and one (1) if the answer is true. Also, we prepared three forms (models) to have the same question and differ in the number of alternatives.

- **Form 1 :**contains five alternatives for each item.
- **Form 2 :**contains four alternatives for each item.
- **Form 3 :**contains three alternatives for each item.

**Two Parameters Model:** One of the models of the Item response theory depends on difficulty and discrimination indicators.

**Information Function Test:** It is a mathematical function which explains a group of information functions of the items.

**Estimation Accuracy**: It is an expression that indicates the quality of estimation which distinguishes a large probability, if the estimation is close to true value (unbiased estimation) using the standard error.

**Item Indicators:** This is a difficulty and discrimination indicators by two parameter models.

**Methodology and Procedures**

- **The Study Society:** It refers to all students who were studying statistics in Jordanian universities. Here, we chose the University of Jordan.
- **The Study Sample:** The University of Jordan and Balqa University in Aqaba.

- **Study Tool:** Preparing a multiple choice exam with 30 question and different choices using the scientific basis of the test building.

✓ The Purpose of the Test: measures the level of student's achievement, and the extent to acquire the basic skills and knowledge

✓ Academic Content Analysis: To analyze the course material for each university

✓ Choosing the questions from the topic.

✓ Preparing the initial test, and applying it to Mutah University.

The researcher used the statistical Program for Social Sciences (SPSS) to find the difficulty indicator and discrimination indicator for each item as shown in table(1) .

Table (1) :The values of discrimination indicators and difficulty for all items in the initial study

| Items | Discrimination | difficulty | Items | Discrimination | difficulty | Items | Discrimination | difficulty |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.29 | 0.81 | 11 | 0.35 | 0.73 | 21 | 0.43 | 0.54 |
| 2 | 0.36 | 0.82 | 12 | 0.28 | 0.58 | 22 | 0.33 | 0.59 |
| 3 | 0.41 | 0.72 | 13 | 0.55 | 0.44 | 23 | 0.56 | 0.38 |
| 4 | 0.38 | 0.29 | 14 | 0.66 | 0.59 | 24 | 0.70 | 0.52 |
| 5 | 0.48 | 0.42 | 15 | 0.41 | 0.66 | 25 | 0.45 | 0.41 |
| 6 | 0.55 | 0.37 | 16 | 0.51 | 0.38 | 26 | 0.29 | 0.33 |
| 7 | 0.35 | 0.51 | 17 | 0.49 | 0.45 | 27 | 0.66 | 0.55 |
| 8 | 0.30 | 0.81 | 18 | 0.62 | 0.61 | 28 | 0.51 | 0.46 |
| 9 | 0.51 | 0.59 | 19 | 0.71 | 0.33 | 29 | 0.50 | 0.71 |
| 10 | 0.50 | 0.61 | 20 | 0.52 | 0.78 | 30 | 0.48 | 0.40 |

Based on table (1), the discrimination indicators ranged from (0.28) to (0.71) with average (0.47); and the difficulty indicators ranged from (0.29) to (0.82), with average (0.55).

However, the value of stability indicator by internal consistency (KR - 20) of the sample was (0.83).

✓ Preparing forms (models) of tests, such that:

- Form 1: three alternatives.
- Form 2: four alternatives.
- Form 3: five alternatives.

**The Basic Procedures of the Study**

Applying the distribution models, the data for analysis is inserted using SPSS and Bilg – Mg software to get the results and examine the assumptions of models by two parameter models.

**The Results of the Required Assumptions by Two Parameter Models**
**Unidimensional (One-Dimensional)**

This study confirmed the assumption of Unidimensional for current study by Factor analysis to the students responses on each model, and with Principal component Analysis using SPSS. Thus, we found the first Eigen value, second Eigen value, and the ratio between them, as well as the percentage of variance in explaining each factor. If the probability value increases the assumption of unidimensional, then the accuracy and sincerity of the result are increasing as seen in table (2).

Table (2): The Eigen Value and the Percentage of Interpretation Variance of Different Distributions

| Form | Factor | Eigen value | Percentage of variance |
|---|---|---|---|
| First Form | 1 | 11.55 | 22.32 |
| | 2 | 4.35 | 7.11 |
| Second Form | 1 | 13.24 | 19.12 |
| | 2 | 4.1 | 9.22 |
| Third Form | 1 | 10.64 | 25.31 |
| | 2 | 3.71 | 3.49 |

Based on table (2), the second Eigen value is greater than the first Eigen value of all tests. However, this means the assumption of the item response theory is satisfying.

**B. Assuming Local Independence:** The local independence assumption is equivalent to unidimensional (Hambleton, Swaminathan & Rogars .1991).

**C. Goodness of Fit Test**: It is an essential step, because the response theory properties realized that there was a good match between the model and the data which has been used (Bilog-Mg3). Also, items and students were not corresponding to the Two parameter model for each model. Then, we excluded Items and students who were not corresponding to the Two parameter logistics model, by depending on Chi-Square Distribution ( $\chi^2$ ) that deletes the student if the probability value is less than $(\alpha = 0.05)$ .

**D. F reedom from Speed**

Speed assumption is implicit in the assumption of unidimensional.

**Statistical Software:** SPSS and BILOG-MG were used to analyze the data and the calculation of:

-The value of the stability of the internal consistency by KR - 20.
-The discrimination and discrimination indicators .
-Estimate the ability of test takers (students ).
-Function information for the three models.
-Use the variance analysis method to answer the study hypotheses .

**The Results of a Study**

This study aimed to investigate the effect of a number of choices (alternatives) in the test of multiple-choice on the difficulty indicator and discrimination indicator for the items and the respondents in the University of Jordan. Thus, we will discuss the following questions:

**Question 1:** Is the accuracy estimate of the difficulty indicator by two parameters model depending on the number of choices (alternatives)?

The test has reflected a number of values on the difficulty indicator and the standard error by the Two parameter models for the three forms (models), as shown in Table (3).

Table (3): Difficulty indicators and standard error for the items of the **three forms**

| Item no. | Form 1 | | Form 2 | | Form 3 | | Item no. | Form 1 | | Form 2 | | Form 3 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Difficulty indicators | Standard Error | Difficulty indicators | Standard Error | Difficulty indicators | Standard Error | | Difficulty indicators | Standard Error | Difficulty indicators | Standard Error | Difficulty indicators | Standard Error |
| 1 | -1.45 | 0.08 | -1.54 | 0.45 | -1.12 | 0.17 | 16 | -0.89 | 0.06 | -0.33 | 0.08 | -0.34 | 0.11 |
| 2 | -0.85 | 0.25 | -1.15 | 0.21 | -1.4 | 0.33 | 17 | 0.62 | 0.04 | 0.04 | 0.08 | 0.22 | 0.07 |
| 3 | -1.33 | 0.26 | -0.57 | 0.12 | 0.77 | 0.22 | 18 | 0.01 | 0.10 | -0.33 | 0.13 | -0.23 | 0.13 |
| 4 | -1.12 | 0.28 | 0.44 | 0.11 | -0.5 | 0.11 | 19 | -0.4 | 0.18 | -0.06 | 0.11 | 0.07 | 0.12 |
| 5 | 0.92 | 0.70 | -0.33 | 0.06 | 0.25 | 0.11 | 20 | -0.15 | 0.111 | -0.72 | 0.23 | -0.67 | 0.20 |
| 6 | -0.89 | 0.3 | 0.25 | 0.17 | -0.73 | 0.13 | 21 | -0.25 | 0.10 | -0.14 | 0.07 | 0.0 | 0.09 |
| 7 | 0.77 | 0.03 | 0.4 | 0.15 | -1.16 | 0.15 | 22 | -0.38 | 0.09 | 0.03 | 0.08 | -0.22 | 0.09 |
| 8 | -1.24 | 0.05 | -2.44 | 0.18 | -0.67 | 0.18 | 23 | -0.26 | 0.08 | 0.02 | 0.15 | 0.12 | 0.10 |
| 9 | -0.31 | 0.16 | -1.2 | 0.12 | -1.12 | 0.13 | 24 | 0.24 | 0.15 | -0.25 | 0.04 | 0.02 | 0.07 |
| 10 | -0.62 | 0.17 | -0.67 | 0.10 | -0.7 | 0.04 | 25 | -0.20 | 0.07 | -0.12 | 0.09 | -0.24 | 0.05 |
| 11 | -0.86 | 0.34 | -0.22 | 0.16 | -0.12 | 0.06 | 26 | -0.13 | 0.08 | 0.05 | 0.07 | 0.0 | 0.10 |
| 12 | -1.1 | 0.09 | -0.21 | 0.2 | -0.44 | 0.12 | 27 | 0.12 | 0.10 | -0.46 | 0.06 | -0.42 | 0.08 |
| 13 | -0.76 | 0.15 | -0.75 | 0.17 | -1.16 | 0.15 | 28 | -0.34 | 0.06 | 0.03 | 0.08 | 0.09 | 0.10 |
| 14 | -0.78 | 0.04 | -1.53 | 0.06 | 0.60 | 0.23 | 29 | 0.01 | 0.09 | 0.02 | 0.23 | 0.16 | 0.16 |
| 15 | -1.35 | 0.56 | 0.77 | 0.03 | 0.34 | 0.19 | 30 | 0.31 | 0.15 | -0.21 | 0.04 | -0.24 | 0.07 |
| | | | | | | | Mean | -0.42 | 0.16 | -0.37 | 0.13 | -0.29 | 0.13 |
| | | | | | | | Standard deviation | 0.62 | 0.15 | 0.66 | 0.08 | 0.54 | 0.06 |

Table (4)

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| Difficulty Indicator F1 | 30 | -1.45 | .92 | -.4220 | .63484 |
| Standard Error F1 | 30 | .03 | .70 | .1640 | .15193 |
| Difficulty Indicator F2 | 30 | -2.44 | .77 | -.3727 | .66715 |
| Standard Error F2 | 30 | .03 | .45 | .1277 | .08357 |
| Difficulty Indicator F3 | 30 | -1.40 | .77 | -.2947 | .54694 |
| Standard Error F3 | 30 | .04 | .33 | .1287 | .06224 |

In table (1) and table (4), the difficulty indicator of the item to model in the first test ranges from (-1.45) to (0.92) with average (-0.422), and standard error ranges from (0.03) to (0.7) with average (0.164). The difficulty indicator of the item to model in the second test ranges from (-2.44) to (0.77) with average (-0.3727), and standard error ranges from (0.03) to (0.45) with average (0.1277). The difficulty indicator of the item to model in the third test ranges from (-1.4) to (0.77) with average (-0.2947), and standard error ranges from (0.04) to (0.33) with average (0.1287).

To clarify the differences in the accuracy of estimating the difficulty parameters of the items of the three test models, we used the variance analysis of the averages of the standard errors in estimating these parameters using SPSS. This can be seen in table (5).

Table(5) :Results of One Way ANOVA of the differences between the averages of the standard errors in estimating the difficulty of the three forms

| Item Indicator | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| | Between Groups | .026 | 2 | .013 | 1.138 | .325 |
| Difficulty | Within Groups | .984 | 87 | .011 | | |
| | Total | 1.010 | 89 | | | |

In table (5), study showed that there were neither statistically significant differences $(\alpha = 0.05)$ attributed to a number of choices (alternatives), nor to the average of standard error to estimate the difficulty indicators. This study disagree with Crehan, Haladyna & Brewer study, and coincided these results with Shizuka, Takeuchi, Yashima, Yoshizawa, and Yamman studies.

*Question 2: Is the accuracy estimate of the discrimination indicator by two parameters model depending on the number of choices (alternatives)***?**

The discrimination indicator and standard error were estimated by Two parameter models for the three test models. The table shows the indicators of discrimination and the standard error for each item as shown in table (6).

Table (6): Discrimination Indicators and standard error for each item of the three test models according to Two parameter models

| Item no. | Form 1 | | Form 2 | | Form 3 | | Item no. | Form 1 | | Form 2 | | Form 3 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Difficulty Indicators | Standard Error | Difficulty Indicators | Standard Error | Difficulty Indicators | Standard Error | | Difficulty Indicators | Standard Error | Difficulty Indicators | Standard Error | Difficulty Indicators | Standard Error |
| 1 | 0.89 | 0.17 | 0.38 | 0.04 | 1.11 | 0.19 | 16 | 1.11 | 0.38 | 0.98 | 0.18 | 1.11 | 0.17 |
| 2 | 0.66 | 0.17 | 1.22 | 0.33 | 0.28 | 0.06 | 17 | 0.41 | 0.11 | 1.11 | 0.19 | 1.26 | 0.29 |
| 3 | 0.75 | 0.13 | 0.65 | 0.13 | 0.39 | 0.09 | 18 | 0.77 | 0.23 | 0.67 | 0.12 | 0.77 | 0.14 |
| 4 | 0.88 | 0.12 | 0.28 | 0.07 | 0.88 | 0.15 | 19 | 1.12 | 0.12 | 0.80 | 0.15 | 0.89 | 0.09 |
| 5 | 0.51 | 0.08 | 0.88 | 0.18 | 0.80 | 0.12 | 20 | 1.62 | 0.77 | 0.44 | 0.07 | 0.35 | 0.02 |
| 6 | 1.22 | 0.19 | 0.70 | 0.11 | 0.78 | 0.21 | 21 | 0.91 | 0.34 | 0.90 | 0.12 | 0.81 | 0.15 |
| 7 | 0.99 | 0.12 | 0.72 | 0.12 | 1.10 | 0.32 | 22 | 0.30 | 0.06 | 0.98 | 0.19 | 0.87 | 0.17 |
| 8 | 0.87 | 0.16 | 1.15 | 0.26 | 0.66 | 0.16 | 23 | 1.22 | 0.12 | 1.12 | 0.08 | 0.81 | 0.19 |
| 9 | 1.05 | 0.37 | 0.76 | 0.23 | 0.59 | 0.06 | 24 | 0.66 | 0.07 | 0.78 | 0.45 | 0.56 | 0.04 |
| 10 | 1.11 | 0.22 | 0.68 | 0.20 | 0.88 | 0.17 | 25 | 1.20 | 0.22 | 0.76 | 0.15 | 1.80 | 0.32 |
| 11 | 0.70 | 0.34 | 0.58 | 0.13 | 1.23 | 0.18 | 26 | 0.95 | 0.28 | 2.23 | 0.19 | 0.82 | 0.18 |
| 12 | 1.13 | 0.23 | 1.19 | 0.33 | 1.21 | 0.20 | 27 | 1.13 | 0.20 | 1.40 | 0.24 | 1.15 | 0.14 |
| 13 | 0.81 | 0.10 | 1.27 | 0.21 | 1.23 | 0.21 | 28 | 1.17 | 0.21 | 1.21 | 0.19 | 0.65 | 0.17 |
| 14 | 1.22 | 0.26 | 0.85 | 0.18 | 0.37 | 0.11 | 29 | 1.08 | 0.20 | 0.55 | 0.08 | 0.23 | 0.09 |
| 15 | 1.44 | 0.40 | 0.39 | 0.09 | 0.45 | 0.07 | 30 | 0.34 | 0.71 | 1.77 | 0.45 | 1.78 | 0.30 |
| | | | | | | | Mean | 0.941 | 0.236 | 0.913 | 0.182 | 0.861 | 0.159 |
| | | | | | | | Standard deviation | 0.311 | 0.166 | 0.415 | 0.101 | 0.392 | 0.786 |

Table(/)

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| Discrimination Indicator F1 | 30 | .30 | 1.62 | .9407 | .31119 |
| Standard Error F1 | 30 | .06 | .77 | .2360 | .16633 |
| Discrimination Indicator F2 | 30 | .28 | 2.23 | .9133 | .41572 |
| Standard Error F2 | 30 | .04 | .45 | .1820 | .10159 |
| Discrimination Indicator F3 | 30 | .23 | 1.80 | .8607 | .39168 |
| Standard Error F3 | 30 | .02 | .32 | .1587 | .07860 |

In table (6) and (7), the Discrimination indicator of the item to the first model in the test ranges from (0.3) to (1.62) with average (0.9407), and standard error ranges from (0.06) to (0.77) with average (0.236). The Discrimination indicator of the item to the second model in the test ranges from (0.28) to ( 2.23) with average (0.9133), and standard error ranges from (0.04) to ( 0.45) with average (0.1820). The Discrimination indicator of the item to the third model in the test ranges from (0.23) to (1.8) with average (0.8607) , and standard error ranges from (0.02) to ( 0.32 ) with average (0.1587) .

To detect the differences in the accuracy of estimating the difficulty and discrimination parameters of the items of the three test models, we used the variance analysis of the averages of the standard errors in estimating these parameters using SPSS .Thus, this can be seen in table (8).

Table (8) :The results of One Way ANOVA of the differences between the averages of the standard errors in the estimation of discrimination indicators of three forms

| Item Indicator | | Sum of Squares | Df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| | Between Groups | .094 | 2 | .047 | 3.207 | .045 |
| Discrimination | Within Groups | 1.281 | 87 | .015 | | |
| | Total | 1.375 | 89 | | | |

In table (5), study showed there were neither statistically significant differences ($\alpha = 0.05$) attributed to a number of choices (alternatives), nor to the average of standard error to estimate the discrimination indicators. This study disagrees with Crehan, Haladyna & Brewer study, and these results coincided with Shizuka, Takeuchi, Yashima, Yoshizawa, and Yamman studies.

## Conclusion

The following conclusion was drawn from this study based on data analysis and interpretation of results. The results of the study showed there were neither statistically significant differences ($\alpha = 0.05$) attributed to a number of choices (alternatives), nor to the average of standard error to estimate the difficulty and discrimination indicators. Therefore, the first form is the best.

## References:

Hambleton, R. K. & Swaminathan, H. (1985) Item Response Theory, Principles and Applications. Boston: Kluwer. Nijhoff Publishing A members of the Kluwer Academic Publishers Group.

Crocker, L., & Algina, J. (1986). Introductions to Classical and Modern Test Theory. Orlando, FL: Hacourt Brace Jovanovich.
Allen, M. J., & Yen, W. M (1979). Introduction to Measurement Theory. California: Cole Publishing Company.
Aiken, J. (1987). Testing With Multiple-Choice Items: Journal Of Development in Education, 20 (4), 44-57.
Nitco, A. J (2001). Educational Assessment of Students (3rd ed.). New Jersey: Prentice-Hall, Inc.
Plake, B., Thompson, A., & Lowry, S. (1981). Effects of Item Arrangement and Test Anxiety on Low scoring Methods. The Journal of Experimental Education, 49, PP. 214-219.
Costin, F. (1970). "The Optimal Number of Alternatives in Multiple – Choice Tests", Empirical Evidence for Mathematical proof. Educational and Psychological Measurement, 30 (2), 353-358.
Straton, R., & Catts, R. (1980). A Comparison of Tow, Three, and Four-Choice Item Tests Given a Fixed Total Number of Choice. Educational and Psychological Measurement, 40(2), 357-365.
Carehan, K., Haladyna, T. (1993). "Use of an Inclusive Option and the Optimal Number of Options for Multiple-Choice Items". Educational and Psychological Measurement, 53 (1), 241-247.
Shizuka, T., Takeuchi, O., Yashima, Y., & Yoshizawa, K. (2006). A Comparison of Three- and Four- Option English Tests for University Entrance Selection Purposes in Japan. Language Testing, 23 (1), 35-57.
Gronlund, N., and Linn, R. (1990). Measurement and Evaluation in Teaching. New York: Macmillan publishing Co., Inc.