ISSN: 1597-5037

ITEM ANALYSIS FOR OPTIMIZATION OF TEACHER-MADE TEST IN ENTREPRENEURSHIP EDUCATION MULTIPLE-CHOICE ITEMS AT UMAR SULEIMAN COLLEGE OF EDUCATION GASHUA, YOBE STATE NIGERIA

Sulaiman Mai Umar[1]*, Nafiu Mansur Ahmed[2], Mohammed Tijjani[3]

[1]Department of Business Education, School of Secondary Education (Vocational), Umar Suleiman College of Education Gashua, Yobe State

Corresponding Author:
E-Mail: Sulaiman016@gmail.com
Phone No. : 07037693369

Abstract
Item analysis is specifically relevant in test construction. In this sense the items that were previously administered to the test takers were analysed in terms of their psychometric properties. In this research the item analysis of 50 MCQs of Entrepreneurship Education examination was conducted, to determine the difficulty index, discrimination index and distracter efficiency. Item analysis enables test paper setters to identify the quality of the items used for student assessment. Further, identifying ideal questions and developing an item bank for future assessment will enhance the quality of the assessment in education. The difficulty index of an item depicts the ease or difficulty of an item. Forty percent of items (20 items) had average difficulty. The mean DI (0.44) was acceptable; collectively, more than 80% of the items were either excellent or good; only one item had negative DI. Distractors are analyzed to determine their usefulness in tests. When distractors do not perform as expected, they must be removed from the test. This study found a low percentage (26%) of NFDs'

Key Words: Item Analysis, Difficulty index, Discrimination index, Distractor Efficiency

Introduction

Teachers' responsibilities in schools set up are multifaceted, one of such responsibilities is constructing test to determine whether the content of learning imparted to the students by the teacher has been acquired or not. However, some of the test items constructed by teachers are not functioning the way they are designed to function. This is because the items are not subjected to any form of analysis to ensure their quality ad effectiveness.

In the purview of psychometrics, Item analysis refers to statistical methods used for selecting test items for inclusion in a test. The concept goes back at least to Guildford (1936). The process of

item analysis varies depending on the psychometric model. For example, classical test theory or the Rasch model call for different procedures. In all cases, however, the purpose of item analysis is to produce a relatively short list of items (that is, questions to be included in a psychological test) that constitute a pure but comprehensive test of one or a few psychological constructs.

Item analysis is specifically relevant in test construction. In this sense the items that were previously administered to the test takers were analysed in terms of their psychometric properties. This research work therefore uses this procedure to ensure optimum quality of teacher-made test with special reference to entrepreneurship education specifically at Umar Suleiman College of Education Gashua, Yobe State Nigeria. This analysis basically involved item difficulty, discrimination index, distracter effectiveness and response frequency

Statement of the Problem

Tests are been developed by a number of teachers in Entrepreneurship Education in Umar Suleiman College of education Gashua for many years, but the psychometric properties of their items in terms of item difficulty, discrimination and distracter effectiveness are not known because the tests were not subjected to item analysis to ensure that they are functioning effectively and serving the purposes for which they are designed. On this note, this research work examined the psychometric properties of 50 items developed for Entrepreneurship Education during 2022/2023 Academic Session.

Objectives of the Study

The major objective of this research work is to examine the psychometric properties of 50 items developed for Entrepreneurship Education during 2022/2023 Academic Session. The following are the specific objectives for this research work:

1. To examine the item difficulty of the 50 items developed for Entrepreneurship Education in 2022/2023 Academic Session.
2. To assess the facility index of the 50 items developed for Entrepreneurship Education in 2022/2023 Academic Session.
3. To determine the distracter effectiveness of the 50 items developed for Entrepreneurship Education in 2022/2023 Academic Session.

Research Questions

The following research questions were asked to guide the conduct of this research work:

1. What is difficulty level of all the 50 items developed for Entrepreneurship Education in 2022/2023 Academic Session?
2. What is the discrimination power of all the 50 items developed for Entrepreneurship Education in 2022/2023 Academic Session?
3. What is the distracter effectiveness of all the 50 items developed for Entrepreneurship Education in 2022/2023 Academic session?

Conceptual Framework

A conceptual framework provides a clear understanding of the variables, relationships, and assumptions that underpin a research study. It outlines the key concepts that the study is investigating and how they are related to each other. It also defines the scope of the study.

Item Analysis

Item analysis According to Kubiszyn and Borich (2013), is a technique that will enable us to assess the quality or utility of an item. It does so by identifying distractors or response options that are not doing what they are supposed to be doing. According to Mehren and Lehmann (1990). Item analysis is like a handmaiden to the builder of tests. It is the process of examining the students' responses to each test item, to judge the quality of the item. Specifically, what one looks for is the difficulty and discriminating ability of the item as well as the effectiveness of each alternative.

Difficulty index

Mok (2005) noted that difficulty index has been identified as one of the methods to determine the difficulty level in examination questions by classifying it into a three level namely easy, moderate and hard. Difficulty index is referred to as a comparison of difficulty in answering the same examination questions by a group of students. According to Sarina et al. (2017), there are two methods to determine the difficulty index in examinations. The first method is using difficulty index for the objective questions and the second method is for the subjective or essay questions. The purposes of difficulty index as highlighted by Anon (2016) are as follows:

1. To identify the concept that needs to be taught again, upon discovering that students cannot answer some particular questions (hard).

2. To identify and report the strength and weakness of curriculum parts, which can and cannot be dominated by students.
3. To give feedback to students regarding their strength and weaknesses on topics assessed.
4. To identify questions that are content bias, like the contents that being highlighted during the teaching session (example: giving more exercise on the topics).

Difficulty index for the Objective Questions

Cheang & Hasni (2008) defined difficulty index as the ratio of the number of students that can answer the questions correctly to the total number of the students who sit for the exams. The formula for calculating the difficulty index for a small group of students is as follows:

Difficulty index = $\frac{T}{N}$

T = Number of students that answer the questions correctly.

N = Total number of the students who sit for the exams.

Item discrimination power

Item discrimination index is an index which indicates how well an item is able to distinguish between the high achievers and low achievers given what the test is measuring (Sidhu, 2012). That is, it refers to the degree to which it discriminates between test takers with high and low achievement. It is obtained from this formula:

Item Discrimination Power: (D) = Number of high scorer who got items right (H) – Number of low scorer who got item right (L)

Total Number in each group (n)

$$D = \frac{H-L}{n}$$

Item discrimination values range from – 1·00 to + 1·00. The higher the discriminating index, the better is an item in differentiating between high and low achievers. Usually, if item discriminating power is a :- positive value when a larger proportion of those in the high scoring group get the item right compared to those in the low scoring group.

- Negative value when more test takers in the lower group than in the upper group get the item right.
- Zero value when an equal number of test takers in both groups get the item right; and
- 1.00 When all test takers in the upper group get the item right and all the test takers in the lower group get the item wrong.

Distracter Effectiveness

The quality of the distractor can be evaluated by frequency of selection and discrimination. A distractor can be defined as functional when it is intended to be plausible for those students with low achievement. For this reason, a distractor is expected to have negative discrimination and to be selected by at least 5% of the participants (Rodriguez, 2005; Tarrant et al., 2009; DiBattista and Kurzawa, 2011; Hingorjo and Jaleel, 2012; Gajjar et al., 2014). Distractor discrimination is usually evaluated with point-biserial correlations, which are correlations between the overall test score and a dichotomous variable (choosing/not choosing the distractor). The Point-biserial Correlation is mathematically the equivalent to Pearson (Product Moment) Correlation; that is, if we have one continuously measured variable Y, $r_{xy} = r_{pb.}$ This can be shown by assigning two distinct numerical values to the dichotomous variable In 2000, Attali and Fraenkel proposed a modified version in which the dichotomous variable contrasts the students who choose the distractor with those who choose the correct option (Attali and Fraenkel, 2000; Gierl et al., 2017). The analysis of distractors could also be performed alongside estimation of student ability and item difficulty, referring to specific item response theory models, i.e., Bock's nominal-response model and Samejima's graded response model (Gierl et al, 2017).

Methodology

The descriptive, analytical study was conducted at Umar Suleiman College of Education Gashua, Yobe State Nigeria. The 50 MCQ items of entrepreneurship Education exam conducted in the 2023/2024 academic session were analyzed. The exam answered by 43 students, included other forms of objective items also. Each MCQ comprised of a stem and four options: one key and three distractors. The correct answer was awarded one score, and the incorrect response or un-attempted item was given zero, with no negative marking. The score ranged from 0-50 which was aggregated when computing the final score. Data Collection and Analysis Permission to conduct the study was obtained from the Dean of Vocational and Technical Education, because the study was not done on human subjects. All 43 answer scripts were retrieved in 2023 and coded. A total of 50 MCQs and 200 options (50 correct answers and 150 distractors) were assessed, to identify the item difficulty, item discrimination and effectiveness of distracters. The high and low group comprised of 25% of students in each group. The DIF I was computed using the formula H+LAIxl00, where H and L are the correct responses in the high and the low groups, and N is the total number of examinees in both groups. Values between 30% and 70% are acceptable, with lower values reflecting the higher difficulty and vice versa, The DI computed

using the formula DI:H-LX2AI expresses the power of the item to differentiate between the high and the low achievers, and ranges between 0 and l. Greater values indicate higher discrimination power. An item with a DI>0.35 is considered to be 'excellent', 0.25-0.34, to be 'good', 0.15 - 0.24, to be 'marginal' and <0.15 to be poor (Guilbert, 1981). A negative DI (less than zero) indicates that low achievers answers the item more correctly than the high achievers. This is possibly due to item flaws or inefficient distractors.

Distractor efficiency (DE) is calculated on the basis of the number of Non-functional distractors (NFDs) in an item and ranges from 0- 100%. NFDs are those options other than the key when selected infrequently by the respondents (<5%) and do not perform their function. An item with no NFDs has a DE of 100%. When the item has one, two or three NFDs, the DE will be 66.66%, 33.33 % and 0, respectively.

Result and Discussion

Table l: Distribution of items in relation to DIF I and DI

Parameters	Items	Percentage	Interpretation	Mean SD
Difficulty Index (DIF I)				67.50 (±14.33)
>70	23	46%	Easy	
30-70	20	40%	Good	
<30	7	14%	Difficult	
Discrimination Index (DI)				0.44(±0.19)
>0.35	29	58%	Excellent	
0.25-0.34	11	22%	Good	
0.15-0.24	4	8%	Marginal	
<0.15	6	12%	Poor	

Total 50 MCQs and 200 options (50 correct answers and 150 distractors) were analyzed. The scores obtained in the exam ranged from 9-46. The mean and standard deviation (SD) of the DIF I (%), DI were 67.50 114.33, 0.44 + 0.19 respectively.

Difficulty and Discrimination Indices: Of the 50 items, 20 items (40%) were good and had average difficulty (DIF I 30-70%). When the DI was assessed, most items were either excellent (n 29, 58%) or good (n: 11; 22%) with DI above 0.24. Items with a DI below 0.25 were either marginal or poor. Only one item had a negative DI.

To be acceptable, an item should have an average difficulty of 30-70 and a discrimination index of not less than 0.25. When the two indices were combined, there were 40 (80%) appropriate items

Table 2: Distractor Analysis: Distribution of Items and Non-functional Distractors

NFDs	Distractor Efficiency	Number of Items	Percentages
0 NFDs	100%	24	48
1 NFDs	65.43%	13	26
2 NFDs	34.57%	10	20
3 NFDs	0%	3	6

Distractor Analysis: Almost three fourth of the 150 distractors (n: 106; 70.66%), were functional with a choice frequency of >5%. The remaining (n=44; 29.33%) were NFDs. Distractor efficiency varied among the items Overall out of 50 items, 24 items had functional distractors (DE : 100%). The DE for the remaining 26 items varied between 0% and 65.43%. A little over one-fourth of 50 items i.e. l3 (26 %) had one NFD (DE: 65.43%), and 10 items (20%) had three NFDs (DE=34.57%). A very small proportion of items (6%) had 0% DE. The overall mean and SD of the DE was 74.30 and 30+0.16 respectively.

Conclusion

The item analysis of 50 MCQs of Entrepreneurship Education examination was conducted, to determine the difficulty index, discrimination index and distracter efficiency. Item analysis enables test paper setters to identify the quality of the items used for student assessment. Further, identifying ideal questions and developing an item bank for future assessment will enhance the quality of the assessment in education. The difficulty index of an item depicts the ease or difficulty of an item. Forty percent of items (20 items) had average difficulty. The mean DI (0.44) was acceptable; collectively, more than 80% of the items were either excellent or good;

only one item had negative DI. Distractors are analyzed to determine their usefulness in tests. When distractors do not perform as expected, they must be removed from the test. This study found a low percentage (26%) of NFDs'

Recommendation

1. This research recommends for further studies that the number of examinees on whom the test was to be administered should be more than 43 candidates. Further studies on larger student population will help to improve the analysis.
2. Test should be randomly selected to avoid any element of sampling bias. The outcome observed in this study is limited to one exam in one course and does not reflect other academic courses. Since this research has not randomly selected the test, there is an element of sampling bias.
3. Item analysis should be encouraged to improve the quality of items for the development of item bank that will be used for future assessment.

COLLEGES OF EDUCATION ACADEMIC STAFF UNION
COEASU
UNITY AND JOB SATISFACTION
KNOWLEDGE FOR SERVICE
2024

Reference

Anon. (2016). *Modul 5 Kesesuaian Item dan Ujian Dlm. HBEF3203: Pengukuran dan Penilaian Dalam Pendidikan, hlm.* 90-109. Program Sarjana Muda Pengajaran KPM-OUM: UniversitiTerbukaMalaysia. http://www.ppdhl.net/v2/images/fbfiles/files/kesesuaian_item_dan_ujian.pdf13Oktober 2010].

Attali, Y., and Fraenkel, T. (2000). *The point-biserial as a discrimination index for distractors in multiple-choice items: deficiencies in usage and an alternative.* J. Educ. Meas. 37, 77–86. doi: 10.1111/j.1745-3984.2000.tb01077.x

Cheang, Q.B. & Hasni, Z. (2008). *Analisis Item dan Pembinaan Ujian – Satu Perbandingan antara Pendekatan Rujukan Norma dan Rujukan Kriteria.* Jurnal Pendidikan Tigaenf 2(2): 112-120.

DiBattista, D., and Kurzawa, L. (2011). *Examination of the quality of multiple-choice items on classroom tests.* Can. J. Scholarsh. Teach. Learn. 2, 1–23. doi: 10.5206/cjsotl-rcacea.2011.2.4

Gajjar, S., Sharma, R., Kumar, P., and Rana, M. (2014). *Item and test analysis to identify quality multiple choice questions (MCQS) from an assessment of medical students of Ahmedabad, Gujarat. Indian J.* Community Med. 39, 17–20. doi: 10.4103/0970-0218.126347

Gierl, M. J., Bulut, O., Guo, Q., and Zhang, X. (2017). *Developing, analyzing, and using Distractors for multiple-choice tests in education*: a comprehensive review. Rev. Educ. Res. 87, 1082–1116. doi: 10.3102/0034654317726529

Guilbert JJ. Educational handbook for health personnel. Revised edition. WHO Offset Publ. 1981;(35):330 p.

Guilford, J. P. (1936). Psychometric methods. New York: McGraw-Hill.

Hingorjo, M. R., and Jaleel, F. (2012). *Analysis of one-best MCQs: the difficulty index, Discrimination index and distractor efficiency*. J. Pak. Med. Assoc. 62, 142–147.

Mok Soon Sang. (2005). *Asas Pendidikan 3–Penilaian Dalam Proses Pendidikan.* 5 th Ed. Kuala Lumpur: Kumpulan Budiman Sdn Bhd.

Rodriguez, M. C. (2005*). Three options are optimal for multiple-choice items: a meta-analysis of* 80 years of research. EMIP 24, 3–13. doi: 10.1111/j.1745-3992.2005.00006.x

Sarina S., Shahida S. and Mohd M. L. (2017). *Pengkelasan Tahap Kesukaran Soalan* Menggunakan Rangkaian Neural. 1 st International Malaysian Educational Technology Convection, hlm. 357-363

Sidhu K.S, (2012), New Approaches to Measurement and Evaluation. New Delhi: Sterling Publishing Ltd India

Tarrant, M., Ware, J., and Mohammed, A. M. (2009). *An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis.* BMC Med. Educ. 9:40. doi: 10.1186/1472-6920-9-409.

COLLEGES OF EDUCATION ACADEMIC STAFF UNION
COEASU
UNITY AND JOB SATISFACTION
UMAR SULEIMAN COLLEGE OF EDUCATION GASHUA
KNOWLEDGE FOR SERVICE
2024