Pragmatic Perspective of Item Analysis Using Microsoft Office Excel Data Analysis Tools

Benjamin Akpezi ETOBRO
Department of Educational Foundations and Counselling Psychology,
Faculty of Education, Lagos State University, Ojo, Lagos, Nigeria
benjamin.etobro@lasu.edu.ng

Olufunmi TAIWO
Department of Educational Foundations and Counselling Psychology,
Faculty of Education, Lagos State University, Ojo, Lagos, Nigeria
olufunmi.taiwo@lasu.edu.ng

Modupe ALAWAYE
Department of Educational Foundations and Counselling Psychology,
Faculty of Education, Lagos State University, Ojo, Lagos, Nigeria
modupe.alawaye@lasu.edu.ng

Abstract
This study examined with practical demonstration the validation process of item generation through item analysis to determine item difficulty level, item discrimination power and item effectiveness adopting Microsoft Office Excel Data Analysis Tools. The study employed a descriptive expo-facto survey research design. Three research questions guided the study. A 20-multiple choice question (MCQ) items in Psychological Testing in Counselling Psychology course. The MCQ was administered to 139 education students who were in their second year in Lagos State University. A sample of 40 students' responses was extracted by systematic random sampling technique. The students' responses were collected and item analysis carried out. Results show that levels of difficulty, discriminating power and distractor effectiveness on the MCQ items varied. Some items were easy and moderately difficult while the others were difficult to answer. It also uncovers that, in regard to discriminating power, some items are well constructed while the others were broad and ambiguously worded. Conclusively, most of the items were reliable and valid because the levels of their difficulty and discrimination power were satisfactory. Teachers and stakeholders in testing should carry out item analysis to avoid ambiguous and broad items adopting Microsoft Office Excel statistical Tools.

Keywords: Microsoft excel, item analysis, item discrimination power, difficulty power, distractor effectiveness

Introduction

Human and national development is anchored on the education. Education is assumed as the pivotal to technological development of a nation, Nigeria inclusive. Education could be seen as access to the development of individual potentials and national development. The goal of education

is the wholistic development of the individuals which consequently leads to acquisition of knowledge and skills. Thus, instructional activities are pre-specified through the statement of instructional objectives. In a formal system of education, the instructors regularly plan instructional activities, monitor learners' progress and continuously assessing the effectiveness of teaching. In order to ascertain the realization of the instructional objectives, the teacher constantly monitors the learners' progress by gathering information on such progress by assessing the learners. In fostering decision-making process by teachers, assessment given by them is not restricted to learners only, but also extended to teachers' activities. Assessment is an important aspect of instructional process. It provides information for decision-making on true reflection of students' progress, educational institutions, programmes and policies with the view of obtaining good results, (Afemikhe, 2005)

The extent to which knowledge and skills acquired are determined through assessment. The information obtained from assessment activities are used as feedback to teachers, students and stakeholders which could be used in modifying instructional activities. Teachers in the process of evaluating progress of instructional activities, employ various assessment techniques which include testing and non-testing tools. For about a decade ago, it was observed that assessment of students' assignments and written tests were carried out through traditional modes of assessments. The traditional mode of evaluation is a technique whereby the testees are required to express themselves displaying their level of achievement in subject content. After the administration of the test, testees' responses are scored and recorded. Characteristically, marking of testees' responses are usually time consuming, feedback to learners is usually delayed and paper consuming; even loss of documents. The integration of Information Communication Technology (ICT) in the instructional process has turned the world into a global village. There has been an increase in new intakes in schools at all levels of educational sectors. The increase in enrollment of students puts an added workload on teachers which limit them from doing other academic activities, such as marking of tests, assignments and drills and thus, prevent them from providing immediate feedback to learners. In order to overcome the herculean task of marking numerous answer scripts, teachers develop multiple-choice questions (MCQs).

Constructing MCQ seems to be complex and requires a lot of time. Mitra et al. (2009), Nagaraja, Ponnudurai & Judson (2009) argued that multiple-choice questions are used mostly in

assessing comprehensively subject contents as summative assessment at the end of a semester or academic sessions. After scoring of the MCQs, there is always the need for the teacher to assess effectiveness of the test items, whether the test items displayed psychometric qualities of the items. As a result of the versatile nature, MCQs are mostly used in assessing the comprehensiveness of knowledge contents and capabilities of students. The MCQs are usually developed by course lecturers or teachers and the items been moderated by experts in the respective disciplines.

In constructing MCQs, the course lecturers or teachers generally, perhaps may not ensure the reliability and validity of the test items. Teachers may find it cumbersome to ensure the validity, reliability and objectivity of test items due to lack of awareness in the validation process. This perhaps may result in poorly developed items which may either be too easy or too difficult. Hingorjo & Jaleel (2012) in Rehman, Aslam & Hassan, (2018) argued that such poorly constructed items may not be able to distinguish between the upper group and lower group testees. In some cases the distractors may not be effective or functional to discriminate the upper group and lower group testees. Consequently, assessment results may not reflect the true performance of testees. Therefore, there is the need to determine the qualities of MCQs so that items could be used in subsequent tests. Assessing the quality of test items is referred to as item analysis. After constructing and assessing test items there is need for the test developer to know the qualities of the test and whether the test scores would reflect the true performance of the testees.

One way of ensuring the qualities of the test are maintained is by adopting item analysis. Item analysis is the process of collating and summarizing testees' responses to determine the psychometric qualities of test items. According to Sharma (2021), it is a process which assesses student's responses to individual test items to determine the quality of these items and quality of test as a whole. The item analysis assists in determining good items and those items that would need modification or improvement, or that could be deleted or discarded from the question or item bank. It allows any aberrant items to be given attention and reviewed. The parameters of item analysis include item discrimination power; item difficulty and pseudo-change score level, (Metsämuuronen, 2017).

Item discrimination power (IDP) is a characteristic of a test item that provides information on how effectively or efficiently an item can distinguish or discriminate between the testees with the higher scores from those with the lower scores. It measures the differences between the

percentages of students in the upper group and those in the lower group who got the items correctly. On the other hand, item difficulty index is the percentage or proportion of testees who got an item correctly. Item difficulty index indicates whether an item is too easy or too difficult. An item with higher percentage item difficulty means that the item is easy.

According to Ogunmakin & Osakuade (2013) teacher-made tests have been seriously critiqued because teacher-made tests lack some level of validity and reliability. Some teachers perhaps, due to their phobia for mathematical computation involved in ensuring adequate psychometric qualities of test items may end up choosing items that are poorly constructed. To overcome this challenge, teachers need to equip themselves the basic rudiments of item analysis by exploring Microsoft Excel application.

Microsoft Excel otherwise known as the spreadsheet is a versatile computer application or package that enables the user to perform a wide range of activities or functions. Microsoft Excel as an analytical tool has about 81 statistical and 59 mathematical functions. The application can be used to perform dynamic calculations and plot graphs and charts of high quality. Microsoft Excel analytical tool has functions applicable to business and engineering problems. The statistical functions that Excel affords the technological world include descriptive and inferential statistics. The descriptive statistics include mean, mode, median, range, variance and standard deviations whereas, the inferential statistics include t-test, analysis of variance, analysis of covariance, correlation multiple linear regressions. Data Analysis with Microsoft Excel provides different means and ways to analyze and interpret data. Data can be analyzed with the relevant Excel commands, functions and tools; item analysis inclusive. This study was carried out to expose readers to the basic steps in performing item analysis adopting Microsoft Excel application in order to determine item difficulty power, item discrimination power and distractor effectiveness.

Purpose of the Study

The general purpose of this research is to explore and understand;

i. The difficulty levels of Psychological Testing in Counselling Psychology test items
ii. The item discrimination levels of Psychological Testing in Counselling Psychology test items
iii. The distractors effective in discriminating between the upper achievers and lower achievers Psychological Testing in Counselling Psychology test items

Research Questions

The following research questions were raised to guide this study;

i. What are the difficulty levels of Psychological Testing in Counselling Psychology test items?
ii. What are the item discrimination levels of Psychological Testing in Counselling Psychology test items?
iii. Are the distractors effective in discriminating between the upper achievers and lower achievers Psychological Testing in Counselling Psychology test items?

Methodology

The study adopted a descriptive ex-post facto research design. Forty students who took Psychological Testing in Counselling Psychology as a course in Educational Foundations and Counselling Psychology Department of the Faculty of Education in 2021/2022 academic session were randomly selected for the study. The only instrument used in the collection of data is the "Psychological Testing in Counselling Psychology Achievement Test (PTCPAT)". Each of the items of the achievement test has four options with a key and three distractors. The key to each item was scored "1" while the distractors were scored "0". Responses of 40 students to twenty items of PTCPAT were obtained from the pool of responses that were submitted after the administration of the test previously. The responses of the 40 students to the twenty items were entered into Microsoft Excel spreadsheet, Figure 1.

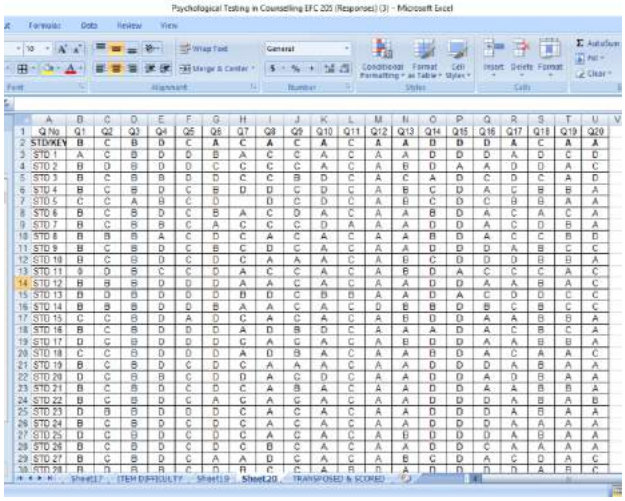


Figure 1: Responses entered in Spreadsheet "1" or "0"

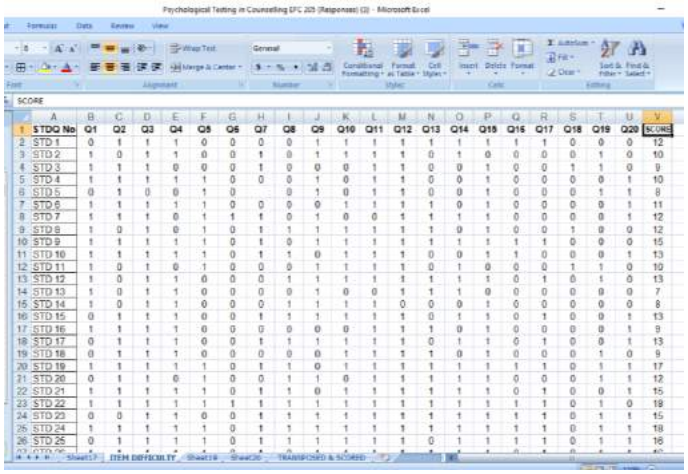


Figure 2: Responses scored either

Research Question One: What are the item difficulty levels of Psychological Testing in Counselling Psychology test items?

In order to answer this question, the data collected were entered into Microsoft Excel Spreadsheet and scored. After scoring of the items, '1' for right answer and '0' for wrong answer, the item difficulty level of each item was computed. The item difficulty level is computed for each item by clicking on Cell B42 and type "= Sum(B2:B41)/40", and then press "Enter" button. To compute the item difficulty index for the rest of the items, place the cursor on "Cell B42" and drag it to cell "U42". This displays the item difficulty indices of all the items as in Figure 3.

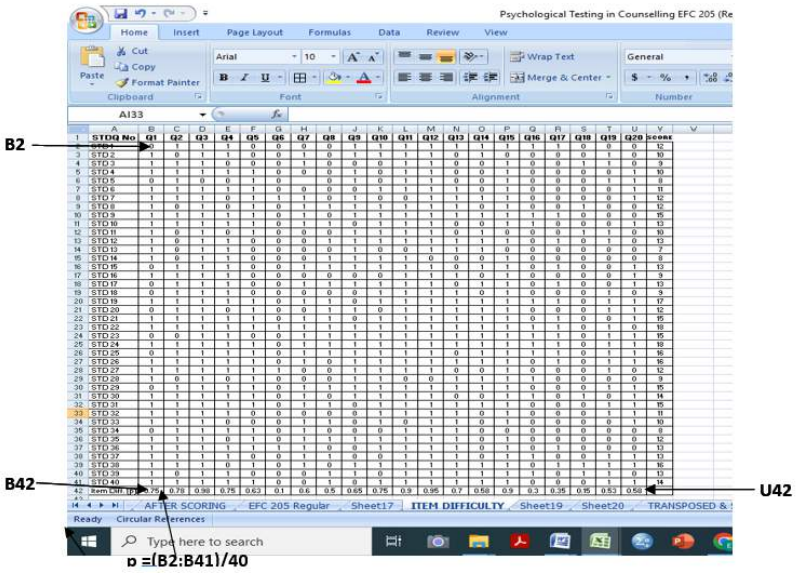


Figure 3: Item Difficulty Index

Table 1
Summary of item Difficulty Index of Psychological Testing in Counselling Psychology Achievement Test (PTCPAT)

Difficulty Index	*f*	%	Items	Item Evaluation
< 0.20	2	10	6, 18	Most difficult
0.20-0.39	2	10	16,17	Difficult
0.40-0.59	4	20	8,19,14,20	Moderately difficult
0.60-0.79	8	40	7,5,9,13,1,4,10,2	Moderately easy
0.9-1.0	4	20	11,15,12,3	Very easy
Total	**20**	**100**		

Table 1, shows that the Difficulty level of the course "Psychological Testing in Counselling Psychology" among Undergraduate Students, ranges 0.01 to 0.98. Two items (items 6 and 18) were

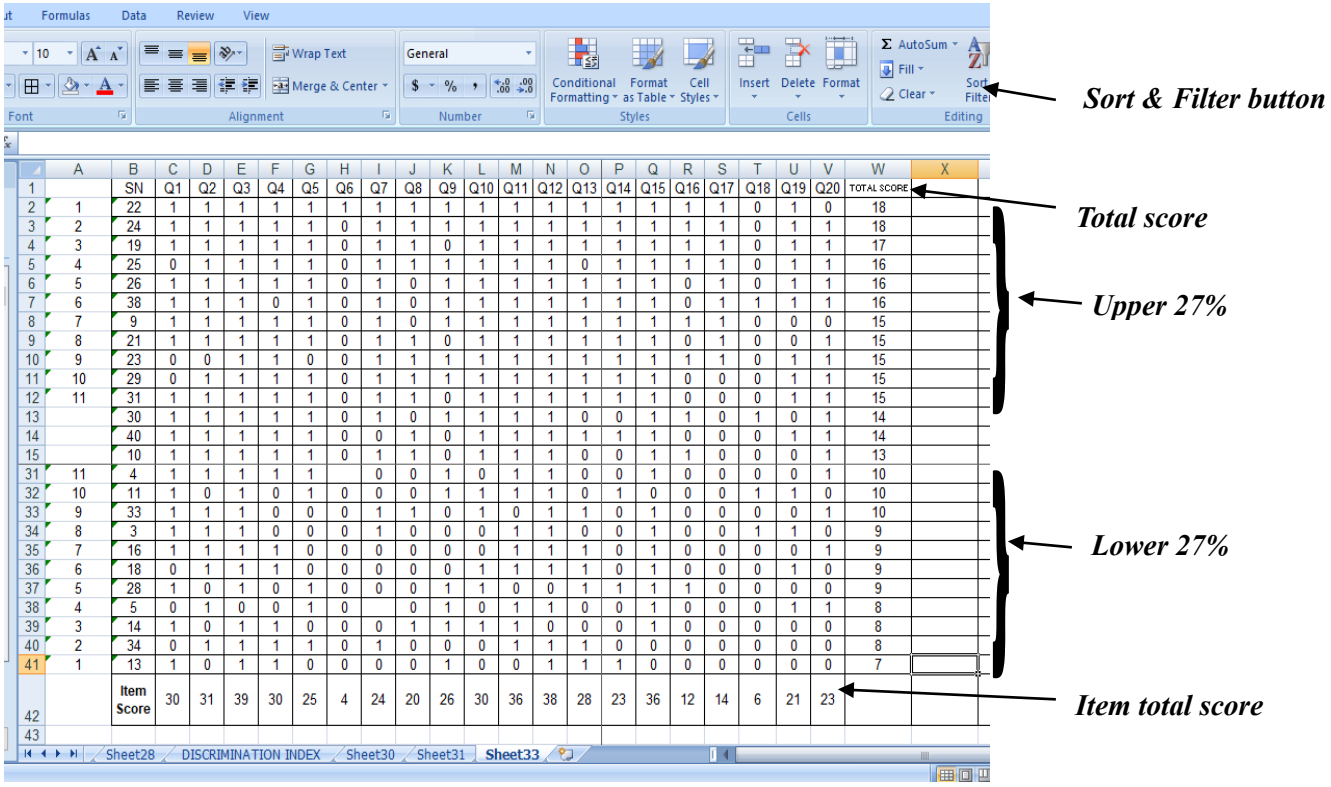
very difficult since they had item difficulty indices less than 0.02 thus, need to be discarded from the pool of items. Furthermore, the table reveals that two items corresponding to 10% of the items had item difficulty levels ranging from 0.40 to 0.59 which indicate that they were moderately difficult hence, they have to be retained. However, eight items corresponding to 40% with item difficulty levels ranging from 0.60 to 0.79 were moderately easy and thus have to be retained. Finally, four items (20%) have item difficulty levels between 0.9 and 1.0 indicating that they are very easy and thus need to be discarded.

Research Question Two: What are the item discrimination levels of Psychological Testing in Counselling Psychology test items?

In order to answer this question, the data collected were entered into Microsoft Excel Spreadsheet and scored. After scoring of the items, '1' for right answer and '0' for wrong answer, the item discrimination index of each item was computed. First the total score is computed for each student. To compute the item discrimination index, the total score is computed for each student by clicking on Cell W2, then Click "∑" on the tool bar. Select Cell "C2" to Cell "V2", then press "Enter" key. The total score for the first testee is now computed. To calculate the total score for the remaining testees, auto-fill by clicking on cell "W2" and drag the cursor to cell "W41". Furthermore, to compute the total number of testees that got each item correctly, use the mouse to select cell "C2" to cell "C41" and then click on the summation symbol "∑" on the tool bar. The total score for the first item is now computed at cell "C42". To compute the total score for the remaining items, auto-fill by clicking on cell "C42" and drag the cursor to cell "V42" Figure 4.

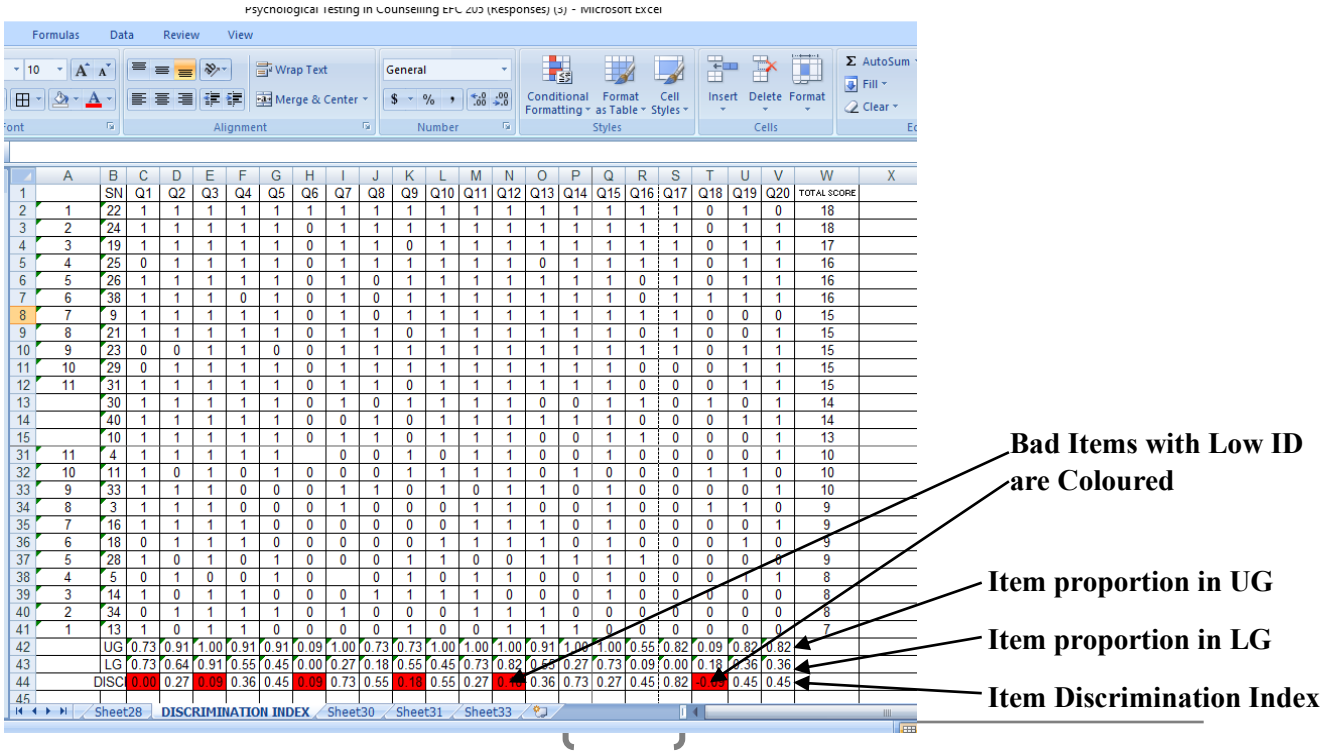
In computing the item discrimination index, arrange testees' total score from the test in descending order by clicking on "Sort & Filter" button on the tool bar. The next step is grouping the students into upper and lower groups based on their scores into upper group that is top 27% of overall scores and lower group that is lowest 27% of overall scores. In this example, 27% of 40 is 11; thus 11 testees are selected from the upper group and another 11 of the overall testees from the lower group as indicated in Figure 3. This is followed by computing the proportion of upper group who answered the item correctly by placing the cursor at cell "C42" and type =SUM(C2:C12)/11. A similar action is performed for the proportion of testees in the lower group who answered the item correctly by placing the cursor at cell "C43" and type =SUM(C31:C41)/11. To obtain the item discrimination power, the proportion of lower group is subtracted from that of the upper group by

clicking on “C44” and typing “=C42-C43”. To compute the discrimination index for the rest of the items, place the cursor on “Cell 44” and drag it to cell “V44”. This displays the discrimination indices of all the items as in Figure 5.



		SN	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	TOTAL SCORE
1																							
2	1	22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	18
3	2	24	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	18
4	3	19	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	0	1	1	17
5	4	25	0	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	16
6	5	26	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	0	1	1	16
7	6	38	1	1	1	0	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1	1	16
8	7	9	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	0	0	15
9	8	21	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	1	0	0	1	15
10	9	23	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	15
11	10	29	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	15
12	11	31	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	0	0	1	1	15
13		30	1	1	1	1	1	0	1	0	1	1	1	1	0	0	1	1	0	1	0	1	14
14		40	1	1	1	1	1	0	0	1	0	1	1	1	1	1	1	0	0	0	1	1	14
15		10	1	1	1	1	1	0	1	1	0	1	1	1	0	0	1	1	0	0	0	1	13
31	11	4	1	1	1	1	1		0	0	1	0	1	1	0	0	1	0	0	0	0	1	10
32	10	11	1	0	1	0	1	0	0	0	1	1	1	1	0	1	0	0	0	1	1	0	10
33	9	33	1	1	1	0	0	0	1	1	0	1	0	1	1	0	1	0	0	0	0	1	10
34	8	3	1	1	1	0	0	0	1	0	0	0	1	1	1	0	0	1	0	0	1	1	9
35	7	16	1	1	1	1	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	1	9
36	6	18	0	1	1	1	0	0	0	0	0	1	1	1	1	0	1	0	0	0	1	0	9
37	5	28	1	0	1	0	1	0	0	0	1	1	1	1	0	0	1	1	1	0	0	0	9
38	4	5	0	1	0	0	1	0		0	1	0	1	1	0	0	1	0	0	0	1	1	8
39	3	14	1	0	1	1	0	0	0	1	1	1	1	0	0	0	1	0	0	0	0	0	8
40	2	34	0	1	1	1	1	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	8
41	1	13	1	0	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	7
42		Item Score	30	31	39	30	25	4	24	20	26	30	36	38	28	23	36	12	14	6	21	23	
43																							

Figure 4: Top and bottom 27% of the testees



		SN	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	TOTAL SCORE
1																							
2	1	22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	18
3	2	24	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	18
4	3	19	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	0	1	1	17
5	4	25	0	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	16
6	5	26	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	0	1	1	16
7	6	38	1	1	1	0	1	0	1	0	1	1	1	1	1	1	1	0	1	1	1	1	16
8	7	9	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	0	0	15
9	8	21	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	1	0	0	1	15
10	9	23	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	15
11	10	29	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	15
12	11	31	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	0	0	1	1	15
13		30	1	1	1	1	1	0	1	0	1	1	1	1	0	0	1	1	0	1	0	1	14
14		40	1	1	1	1	1	0	0	1	0	1	1	1	1	1	1	0	0	0	1	1	14
15		10	1	1	1	1	1	0	1	1	0	1	1	1	0	0	1	1	0	0	0	1	13
31	11	4	1	1	1	1	1		0	0	1	0	1	1	0	0	1	0	0	0	0	1	10
32	10	11	1	0	1	0	1	0	0	0	1	1	1	1	0	1	0	0	0	1	1	0	10
33	9	33	1	1	1	0	0	0	1	1	0	1	0	1	1	0	1	0	0	0	0	1	10
34	8	3	1	1	1	0	0	0	1	0	0	0	1	1	1	0	0	1	0	0	1	1	9
35	7	16	1	1	1	1	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	1	9
36	6	18	0	1	1	1	0	0	0	0	0	1	1	1	1	0	1	0	0	0	1	0	9
37	5	28	1	0	1	0	1	0	0	0	1	1	1	1	0	0	1	1	1	0	0	0	9
38	4	5	0	1	0	0	1	0		0	1	0	1	1	0	0	1	0	0	0	1	1	8
39	3	14	1	0	1	1	0	0	0	1	1	1	1	0	0	0	1	0	0	0	0	0	8
40	2	34	0	1	1	1	1	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	8
41	1	13	1	0	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	7
42		UG	0.73	0.91	1.00	0.91	0.91	0.09	1.00	0.73	0.73	1.00	1.00	1.00	0.91	1.00	1.00	0.55	0.82	0.09	0.82	0.82	
43		LG	0.73	0.64	0.91	0.55	0.45	0.00	0.27	0.18	0.55	0.45	0.73	0.82	0.64	0.27	0.73	0.09	0.00	0.18	0.36	0.36	
44		DISC	0.00	0.27	0.09	0.36	0.45	0.09	0.73	0.55	0.18	0.55	0.27	0.18	0.36	0.73	0.27	0.45	0.82	-0.09	0.45	0.45	
45																							

Figure 5: Item Discrimination Index

Table 2
Summary of item Discrimination Level of Psychological Testing in Counselling Psychology Achievement Test (PTCPAT)

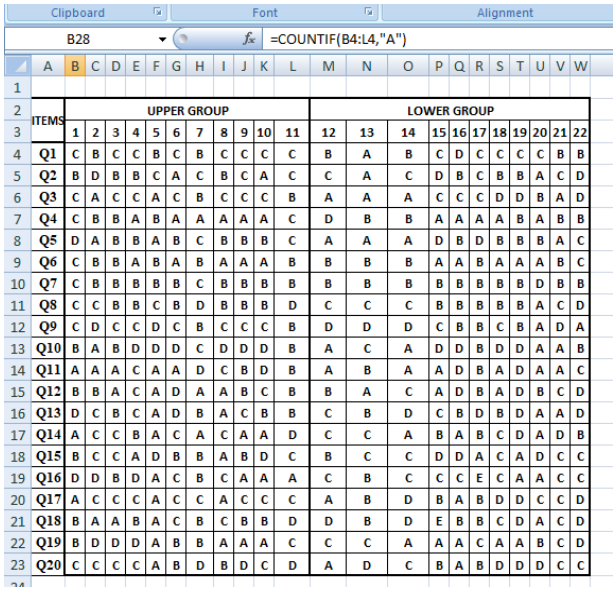
Discrimination Level	*f*	%	Items	Remarks
0.41-0.70	9	45	5,7,8,10,14,16,17,19,20	Good
0.21-0.40	5	25	2,4,11,13,15	Satisfactory
0.00-0.20	5	25	1,3,6,9,12	Poor
Negative	1	5	18	Rejected

Table 2 shows that nine (45%) items: 5, 7, 8, 10, 14, 16, 17, 19 and 20 had item discrimination power of between 0.41-0.70, the nine items distinguished between the upper and lower achievers. Results furthermore reveals that five (25%) items: 2, 4, 11, 13 and 15 had item discrimination index of 0.21 to 0.40 and were considered to be satisfactory because they distinguished effectively between the upper and lower achievers. Finally, five items corresponding to 25% were poor items because their discrimination index is between 0.00 and 0.20 and hence need to be modified. However, one item (5%) had negative discrimination index; hence it should be discarded from the pool of items. Conclusively, about 70% of the items have item discriminating indices ranging from 0.21 to 0.70.

Research Question Three: How effective are the distractors of the items in Psychological Testing in Counselling Psychology objective questions in 2021?

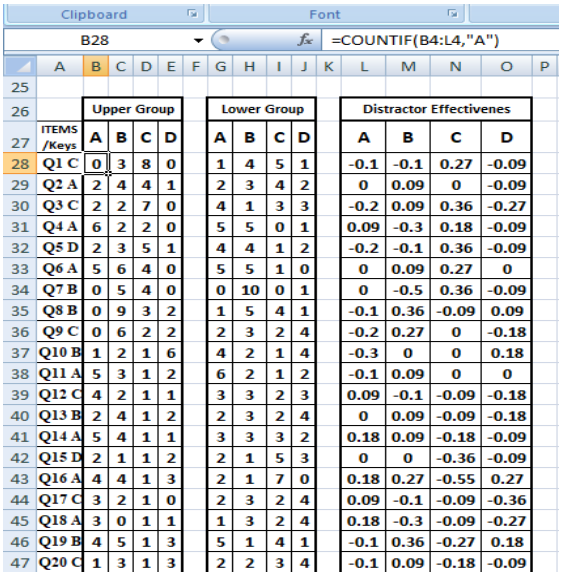
It is noteworthy that item distractor effectiveness index ranges from -1 to +1. A distractor that has a value of close to -1 indicates that the distractor is functional and effective, since it attracts more respondents from the lower group than those in the upper group. However, a distractor with a value of zero implies that the distractor is non-functional and ineffective because such distractor could not distinguish between lower and upper groups as equal number of respondents from both groups have chosen the distractor; such a distractor needs to be replaced or reframed. Meanwhile, a distractor with a value of greater than zero up to plus one (+1), means the distractor is non-functional and ineffective since such a distractor attracted more respondents from the upper group than those in the lower group.

Now to answer the research question, the data collected were subjected to descriptive statistics of frequency counts of the distractors of each item in Microsoft Excel. The options to each of the test items were entered into the cells corresponding to each item as indicated in Figure 6.



		UPPER GROUP										LOWER GROUP									
	ITEMS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
4	Q1	C	B	C	C	B	C	B	C	C	C	C	B	A	B	C	D	C	C	C	C	B	B
5	Q2	B	D	B	B	C	A	C	B	C	A	C	C	A	C	D	B	C	B	B	A	C	D
6	Q3	C	A	C	C	A	C	B	C	C	C	B	A	A	A	C	C	C	D	D	B	A	D
7	Q4	C	B	B	A	B	A	A	A	A	A	C	D	B	B	A	A	A	A	B	A	B	B
8	Q5	D	A	B	B	A	B	C	B	B	B	C	A	A	A	D	B	D	B	B	B	A	C
9	Q6	C	B	B	A	B	A	B	A	A	A	B	B	B	B	A	A	B	A	A	A	B	C
10	Q7	C	B	B	B	B	B	C	B	B	B	B	B	B	B	B	B	B	B	B	D	B	B
11	Q8	C	C	B	B	C	B	D	B	B	B	D	C	C	C	B	B	B	B	B	A	C	D
12	Q9	C	D	C	C	D	C	B	C	C	C	B	D	D	D	C	B	B	C	B	A	D	A
13	Q10	B	A	B	D	D	D	C	D	D	D	B	A	C	A	D	D	B	D	D	A	A	B
14	Q11	A	A	A	C	A	A	D	C	B	D	B	A	B	A	A	D	B	A	D	A	A	C
15	Q12	B	B	A	C	A	D	A	A	B	C	B	B	A	C	A	D	B	A	D	B	C	D
16	Q13	D	C	B	C	A	D	B	A	C	B	B	C	B	D	C	B	D	B	D	A	A	D
17	Q14	A	C	C	B	A	C	A	C	A	A	D	C	C	A	B	A	B	C	D	A	D	B
18	Q15	B	C	C	A	D	B	B	A	B	D	C	B	C	C	D	D	A	C	A	D	C	C
19	Q16	D	D	B	D	A	C	B	C	A	A	A	C	B	C	C	C	E	C	A	A	C	C
20	Q17	A	C	C	C	A	C	C	A	C	C	C	A	B	D	B	A	B	D	D	C	C	D
21	Q18	B	A	A	B	A	C	B	C	B	B	D	D	B	D	E	B	B	C	D	A	C	D
22	Q19	B	D	D	D	A	B	B	A	A	A	C	C	C	A	A	A	C	A	A	B	C	D
23	Q20	C	C	C	C	A	B	D	B	B	D	C	D	A	D	C	B	A	B	D	D	C	C

Figure 6: Options entered into the appropriate cells



	ITEMS /Keys	Upper Group				Lower Group				Distractor Effectivenes			
		A	B	C	D	A	B	C	D	A	B	C	D
28	Q1 C	0	3	8	0	1	4	5	1	-0.1	-0.1	0.27	-0.09
29	Q2 A	2	4	4	1	2	3	4	2	0	0.09	0	-0.09
30	Q3 C	2	2	7	0	4	1	3	3	-0.2	0.09	0.36	-0.27
31	Q4 A	6	2	2	0	5	5	0	1	0.09	-0.3	0.18	-0.09
32	Q5 D	2	3	5	1	4	4	1	2	-0.2	-0.1	0.36	-0.09
33	Q6 A	5	6	4	0	5	5	1	0	0	0.09	0.27	0
34	Q7 B	0	5	4	0	0	10	0	1	0	-0.5	0.36	-0.09
35	Q8 B	0	9	3	2	1	5	4	1	-0.1	0.36	-0.09	0.09
36	Q9 C	0	6	2	2	2	3	2	4	-0.2	0.27	0	-0.18
37	Q10 B	1	2	1	6	4	2	1	4	-0.3	0	0	0.18
38	Q11 A	5	3	1	2	6	2	1	2	-0.1	0.09	0	0
39	Q12 C	4	2	1	1	3	3	2	3	0.09	-0.1	-0.09	-0.18
40	Q13 B	2	4	1	2	2	3	2	4	0	0.09	-0.09	-0.18
41	Q14 A	5	4	1	1	3	3	3	2	0.18	0.09	-0.18	-0.09
42	Q15 D	2	1	1	2	2	1	5	3	0	0	-0.36	-0.09
43	Q16 A	4	4	1	3	2	1	7	0	0.18	0.27	-0.55	0.27
44	Q17 C	3	2	1	0	2	3	2	4	0.09	-0.1	-0.09	-0.36
45	Q18 A	3	0	1	1	1	3	2	4	0.18	-0.3	-0.09	-0.27
46	Q19 B	4	5	1	3	5	1	4	1	-0.1	0.36	-0.27	0.18
47	Q20 C	1	3	1	3	2	2	3	4	-0.1	0.09	-0.18	-0.09

Figure 7: **Item distractor effectiveness**

In order to perform item distractor effectiveness, the frequency of an option in each item is counted first in the upper group by placing the cursor at cell "B28" and type **=COUNTIF(B4:L4,"A")/11**. Then press "Enter" button to obtain the frequency of Option "A" in item question 1; which is zero "0" in this case. Place the cursor at cell "B28" to auto fill question 2 to question 20 to obtain the frequency of option A by dragging the cursor from "B28" to "B47". To compute the frequency of option "B" in upper group, place the cursor at cell "C28" type "**=COUNTIF(B4:L4,"B")/11"** and press "Enter" button. Place the cursor at cell "C28" to auto fill question 2 to question 20 to obtain the frequency of option B. Perform similar actions for options "C" and "D".

Item distractor effectiveness can be achieved by subtracting the frequency of an option in the lower group from that of the upper group and divide by the 27% of the total, as shown in Figure 6. To compute the item distractor effectiveness for option A, place the cursor at cell "L28" and type **"=(B28-G28)/11"** and press the "Enter" button. Auto-fill the item distractor effectiveness for question items 1 to 20 by dragging the cursor from cell "L28" to cell "L47". Perform similar actions for options B, C and D, as in Figure 7.

Table 3: Distractor Analysis Showing the Index of Effectiveness of each Item Distractor

Item distractor effectiveness index	*f*	%	Items	Remarks
Negative index of effectiveness	33	55	1A,B,D; 2D;3A, D; 4B, D; 5A, B, 7D; 8A, C; 9A, D; 10A; 12B, D; 13C, D; 14C, D; 15C, 16C; 17B, D; 18B, C, D; 19A, C; 20A, D.	Distractors/options in each of the items are plausible. More students in the lower group selected the distractor than those in the higher group. Retain distractors.
Zero (0)	8	13	2C; 6D; 7A; 10C; 11 C, D; 15 A, B	Non-functional and ineffective. Discard distracters/options
Positive index between 0.09-0.36	19	32	2B, 3B, 4C; 5C; 6B, C; 7C; 8D; 9B; 10D; 11B; 12A; 14B; 16B,D; 17 A; 19B, D; 20B	Distractors are not plausible. More students in the higher group selected the distractor than those in the lower group. Revise distractors.

From Table 3, 55% of the 60 distracters have distractors with negative index of effectiveness. This implies that the distractors in each of the items are plausible indicating that there were more students in the lower group who selected the distractors than in the upper group. Such item distractors are retained.

Furthermore, Table 3 reveals that 32% of the items have distractors with positive index of effectiveness. This means that the distractors in each of the items are not plausible which indicates that there were more students in the upper group who selected the distractors compared to those in the lower group. Such distractors in the items require revision.

However, 13% of the distractors chosen by equal number of students from upper and lower groups indicating that these item distracters were not plausible (non-functional and ineffective), thus the distractors have to be discarded.

Summary of findings

In this study 10% of the items were very difficult and needed to be discarded. However, 20% of the items were very easy thus; they also have to be discarded. Furthermore, 70% of the items discriminate effectively between upper and lower achievers hence the items were retained; while 25% of the items discriminated poorly, hence they have to be modified and 5% of the items had negative discrimination index, thus they were discarded. The study showed that 55% of the 60 distractors have distractors with negative index of effectiveness. The item distractors were retained. Furthermore, 32% of the items have distractors with positive index of effectiveness. Such

distractors in the test items were recommended for revision. However, 13% of the distractors chosen by equal number of students from upper and lower groups thus the distractors were recommended to be discarded.

Discussion of Findings

The finding in this study revealed that the level of item difficulty in the test was not equally distributed. This finding is in consonant with those of Abdul, Ayesha, & Syed Hammad (2018) who reported than more 50% of the items had a difficulty index of more than 0.71. Furthermore, Toksöz and Ertunç (2017) reported in their study that almost half of the items were moderately difficult. Similarly, Danuwijaya (2018) and Yumelking (2019) both respectively reported that the test items were filled with moderately difficult and difficult items. It is evident in the various studies that difficulty indices of test items are not usually evenly distributed. It is worthy of note that test items should not be too easy or difficult but rather should measure what they suppose to measure. Thus, Gronlund (1998) in Jannah et al.(2021), Hidayat, Husna & Khasbani (2021) suggested that item analysis should be targeted to what specific language tasks testees are or cannot perform, not to discriminate between upper and low achievers.

Result in this study further showed that most (70%) of the items had item discriminating indices ranging from 0.21 to 0.70, while 25% of the items discriminated poorly, their discrimination index is between 0.00 and 0.20 and 5% had negative discrimination index. This trend of the test items with varied discriminating indices is in agreement with the study of Hartanti and Yogi (2019) who reported that the test items analysed had all levels of discriminating index. Likewise, Danuwijaya (2018) in his study also stressed different levels of difficulty on test items.

The finding in this study showed that 55% of the 60 distractors have distractors with negative index of effectiveness. Furthermore, 45% of the items have non-functional distracters. This is in agreement with the study of Gajjar *et al* (2014) and Hingorjo *et al* (2012) in their studies reported distractor effectiveness means of 88.6% and 81.4%, respectively. However, Rehman, Aslam and Hassan (2018) in their study that 51.6% of the distractors were non-functional while, Namdeo and Sahoo (2016) reported 53.4% non-functional distracters. It should be noted that having non-functional distracters would increase item difficulty index and subsequently making test item easier for the testees.

Conclusion

Teachers are faced with enormous tasks of ensuring holistic development of learners. The process of assessing and evaluating the students or learners takes time and energy especially in the scoring and grading of students scripts. Objective assessment of learners requires generating good quality items through item analysis. The analysis of the qualities of a test helps to improve on assessments of achievement of learners. Items with moderate level of difficulty, higher item discrimination power and functional distractors should be included in a test in order to increase the psychometric properties of a test. This would in turn help improve the overall test scores and effectively distinguish among testees of different achievement levels. Teachers must be properly trained in assessment designing so that effective and meaningful test may be conducted. Large number of test items overtime could be kept in question banks for subsequent usage.

Recommendations

Thus, in this era of globalization and technological advancement it is recommended that item analysis operation be adopted by teachers and lecturers in order to reduce the amount of time and money in the preparation and development of quality test items as well as conserve papers. This could be achieved among others through the adoption of well-developed multiple choice questions (MCQs) void of ambiguities with items that would distinguish between upper ability and low achievers. It should worthy of note that an articulated items generation procedure that employs item analysis would raise the psychometric properties of the test.

References

Abdul, R., Ayesha, A., & Syed Hammad, H. (2018). Item analysis of multiple choice questions. *Pakistan Oral & Dental Journal*, 38(2), 291–293.

Afemikhe, O.A. (2005). Selection into Nigeria University. Paper presented at the Annual Conference of International Association for Educational Assessment.

Danuwijaya, A. A. (2018). Item analysis of reading comprehension test for post-graduate students. *English Review: Journal of English Education*, *7*(1), 29–40. https://doi.org/10.25134/erjee.v7i1.1493.

Gajjar S, Sharma R, Kumar P, 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*. 2014;39(1):17-20.

Hartati, N., & Yogi, H. P. S. (2019). Item analysis for a better quality test. *English Language in Focus (ELIF)*, *2*(1), 59–70. https://doi.org/10.24853/elif.2.1.59-70

Hingorjo MR, Jaleel F (2001). Analysis of One-Best MCQs: the Difficulty Index, Discrimination Index and Distractor Efficiency. *J Pak Med Assoc*. 2012;62(2):142-7.

Jannah, R., Hidayat, D. N., Husna, N. & Khasbani, I (2021). An item analysis on multiple-choice questions: a case of a junior high school English try-out test in Indonesia. *Leksika*.Vol.15 (1), 9-17

Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., Rizvi, M. (2017). Difficulty index, discrimination index and distractor efficiency in multiple choice questions. *Annals of PIMS*, 310-315.

Mitra N. K., Nagaraja, H. S., Ponnudurai G & Judson, J. P., (2009). The Levels Of Difficulty And Discrimination Indices In Type A Multiple Choice Questions Of Pre-clinical Semester 1 Multidisciplinary Summative Tests. *IeJSME.* 3 (1), 2-7.

Namdeo, S. K, Sahoo B. (2018). Item analysis of multiple choice questions from an assessment of medical students in Bhubaneswar, India. *International Journal Res Med Sci.* 4(5):1716-19.

Rehman, A., Aslam, A. & Hassan, S. H. (2018). Item analysis of multiple choice questions**.** *Pakistan Oral & Dental Journal*. 38 (2).

Sharma, L. R. (2021). Analysis of Difficulty Index, Discrimination Index and Distractor Efficiency of Multiple Choice Questions of Speech Sounds of English. *International Research Journal of MMC (IRJMMC).* 2 (1), 15-28.

Toksöz, S., & Ertunç, A. (2017). Item analysis of a multiple-choice exam. *Advances in Language and Literary Studies,* 8(6), 141–146.