ITEM ANALYSIS OF TEST OF NUMBER OPERATIONS

EVRORO, Edhereveno Sylvanus
Department of Technical and Business Education
Delta State University, Abraka, **NIGERIA**

ABSTRACT

The Test of Number Operation is a test constructed to test the ability of students in the operation of numbers. It is made up of 35 items with 17 multiple choice items. The item analysis of Test of Number Operations was carried out using 30 students from SS 1 from a rural school in Isoko South Local Government Area of Delta State. The 30 students were those who were interested in taking the Test out of the 85 students in SS 1. The average age of the students is 15 years and male students are 20 while female students are 10. The correction for guessing was applied. Although only 17 out of the 35 items in the test are multiple choice items, it was assumed that all the items are multiple choice items to enable the use of formula for correction. The formula used is S = $R - \frac{W}{n-1}$ The item difficulty. index (p) is given by the formula P = $\frac{R}{T}$ x 100 or P = R/T, The item discriminating index used is $\frac{R_u - R_L}{\frac{1}{2}T}$
The item analysis of the test of Number operation revealed that the range of the actual score of the 30 students is 12 i.e (18-6) while the range of the true score is 14.4 i.e. (14.6-0.2). It was concluded that the items in the Test of Number Operation should be worked upon to increase its item statistics. The item analysis of Test of Number operation should be carried out in different parts of the state and the sample should be increased.

Keywords: Item Analysis, Item Difficulty, item Discrimination Index and distracters.

INTRODUCTION

The basic concepts and techniques of item analysis, like knowledge about other phases of test construction, can help test users in the evaluation of published and teacher made tests. Items can be analyzed qualitatively, in terms of their content and form, and qualitatively, in terms of statistical properties. Qualitative analysis includes consideration of content validity, as well as the evaluation of terms in terms of effective item writing procedures. Qualitative analysis includes principally the measurement of item, difficulty, and item discrimination. This study was focused on the second dimension of item analysis namely the qualitative analysis of Test of Number Operations using thirty students.

The Meaning of Item Analysis

Item analysis is a systematic procedure designed to obtain specific information about each item of a test. It is designed primarily for use with objective test (especially multiple choice items). In item analysis, the test constructor is concerned with item, difficulty level, the discriminative power of the item and effectiveness of the distracters (Ukwuije, 1996). Item

difficulty is an indication of the extent to which an item is difficult for testees or respondents. An item difficulty level is determined by estimating the percentage of examinees that are likely to get the item right when it is administered.

The term "discriminating index" is used to indicate the extent to which response to an item could distinguish between the strong and the weak students like difficulty index, most items are found to have discriminating indices varying between zero to 1.00.

The effectiveness of each distracter is the degree to which it attracts more low achievers than high achievers. How well each distracter is operating can be determined by inspection and there is no need to calculate an index of effectiveness, although the formula for discriminating power could be used for the purpose (Gronlund, 1976).

Uses of Item Analysis

The dividends of item analysis are numerous both to the test constructors test users, testees and test analysts. Item analysis provides diagnostic information for determining the quality of the items. In the view of Crobbach (1990: 178), statistical analysis of items spots questionable items. When these items are reviewed or rewritten they increase the validity of the test.

In short item analysis should lead to the improvement of individual test items. The analysis of each of the items will enable the test constructor/users know the effectiveness of each item and the stem in case of multiple choice item.

Item analysis makes it possible to shorten a test and at the same time increase its validity and reliability. This is achieved because it helps us to choose items of suitable difficulty level. Item analysis lead to increased skill in test construction. Item analysis reveals ambiguity, clues, ineffective distracters, and other technical defects that were missed during the preparation of the test. This information is used directly in the revision of the test items for future use. As test constructors we analyze testees response to items we become increasingly cognizant of technical defects and the factors causing them and in doing so this will increase our skills in test constructions.

In the view of Gronlund (1976) item analysis has a number of fringe benefits of special value to classroom teachers. The most important of these include:

Item analysis data provide a basis for efficient class discussions of the test results. When wee know how effectively each item functioned in measuring achievement, it enables us (teachers) to confine the discussion to those areas which will be most helpful to pupils. Any item that all the testees answered correctly is omitted in the course of discussion while attention is drawn to those items that pose a problem to most of the pupils. During the discussion some items may have more than one answer and both answers are taken and counted as correct and the scoring adjusted accordingly.

Item analysis data provides a basis for the general improvement of classroom instruction. Item analysis data can help in evaluating the appropriateness of the specific learning outcomes and the course content for the particular type of pupils being taught. For instance if all the items from a particular content area are consistently too difficult for the pupils, it might suggest curriculum revisions or shifts in teaching emphasis, similarly errors in pupils)

thinking which persistently appear in item analysis data might indicate that there is need for more effective teaching procedures. In these ways, item analysis can provide insight into the instructional weaknesses requiring more extended attention. When item analysis identifies the areas of weakness, special attention can be given to them in form of remedial work.
Item statistics can be used to forecast roughly the total-score statistics (Guilford, 1984)

Steps in Item Analysis

Gronhand (1976: 264-265) identified the following steps as procedure for item analysis.

I. Rank the scores in order from the highest to the lowest score
2. Select the 27% of individual (testees) with the highest scores and the 27% of the testees with lowest scores.
3. For each item of the test, tabulate the number of testees in the upper and lower groups who select each alternative.
   This tabulation can be made directly on the test paper or on the test item card.
4. Estimate the difficulty of each tern (percentage of testees who got the item right in the upper and lower groups).
5. Estimate the discriminating power of each item (difference between the number of testees in the upper and lower groups who got the item right).
6. Evaluate the effectiveness of the distracters in each item (attractiveness of the incorrect alternatives).

The first three steps of this procedure merely provide a convenient tabulation of testees responses from which we can readily obtain an estimate of item difficulty, item discriminating power and the effectiveness of each distracter. This latter information can frequently be obtained simply by inspecting the item analysis data.

Summary of Item Analysis of Test of Number Operations and Observations

The Test of Number Operations is a test assessing the ability of students in the operation of numbers. It is made up of 35 items with 17 multiple choice items.

The item analysis of Test of Number Operations was carried out using 30 students from SS 1 from a rural school in Isoko South Local Government Area of Delta State. The 30 students were those who were interested in taking the test out of 85 students in SS 1. The average age of the testees is 15 years and male students are 20 while female students are 10.

The Actual and True Score

The Actual scores of the testees were recorded and the correction for guessing was applied. Although only 17 out of the 35 items in the test are multiple choice items, it was assumed that all the items are multiple choice items to enable the use of formula for correction

$$S = R - \frac{W}{n-1}$$

where S = the true score
R = the number of item scored correctly, W = the number of item scored wrongly
n = the number of alternative for each item (n=5)

Table 1: The Actual and True Scores for 30 SS I students who took the Test of Number Operations-

S/N	Sex	Actual Score	True score	S/N	Sex	Actual score	True score
1	M	18	14.6	16	M	10	5.0
2	M	17	14.4	17	M	10	5.0
3	M	16	12.2	18	F	10	5.0
4	M	16	12.2	19	M	10	5.0
5	M	16	12.2	20	M	10	5.0
6	F	16	12.2	21	M	09	3.8
7	F	15	11.0	22	F	09	3.8
8	M	15	11.0	23	F	08	2.6
9	M	15	11.0	24	M	08	2.6
10	F	13	8.6	25	M	08	2.6
11	F	13	8.6	26	F	08	2.6
12	M	12	7.4	27	M	07	1.4
13	F	12	7.4	28	F	07	1.4
14	M	12	7.4	29	M	07	1.4
15	M	11	6.2	30	M	06	0.2

The true score becomes very significant when the testees score is very low. For example a testee whose actual score is 32 will have a true score of 31.25 and the difference between the true score and the actual score is not significant but a testee who score 10, the true score will be 5 which is half the actual score and the difference between the two scores becomes very significant.

Determination of Item Difficulty and Discriminating Index The difficulty index of an item (p) is given by the formula. $P = \frac{R}{T}$ x 100 or P= R/T

where P = difficulty level or difficulty index

R = the number of testees who got the item right in both the upper and lower groups

T = the total number of testees in the upper and lower groups

The discriminating power of an item is determined using the formula:

Item Discriminating Index = $\frac{R_u - R_L}{\frac{1}{2}T}$

where Ru = Upper 27% of the score

RL = Lower 27% of the score

T = Total number of testee in upper and lower groups

Table 2: Determination of Item Difficulty level and Item Discrimination index

Item	Upper 27% (8)	Lower 27% (8)	Difficulty level (%)	Discrimination Index
1	7	5	75.0**	0.25
2	6	5	68.8*	0.13
3	6	2	50.0*	0.50*
4	1	0	6.3	0.13
5	4	3	43.8	0.3
6	2	2	25.0	0.00
7	1	0	6.3	0.13
8	8*	1	56.3*	0.88*
9	2	1	18.8	0.13
10	0	0	0.0	0.00
11	2	0	12.5	0.25
12	8	7	93.8**	0.13
13	8	5	81.3**	0.25
14	8	8	0.0	0.00
15	1	1	12.5	0.00
16	0	0	0.0	0.00
17	2	1	18.8	0.13
18	2	1	18.8	0.13
19	0	2	12.5	(0.25)
20	0	0	0.0	0.00
21	4	0	25.0	0.50*
22	8	3	68.8*	0.63*
23	8	4	75.0*	0.50*
24	7	1	50.0*	0.75*
25	6	2	50.0*	0.50*
26	1	0	6.3	0.13
27	8	1	56.3*	0.88*
28	0	0	0.0	0.00
29	0	0	0.0	0.00
30	1	0	6.3	0.13
31	5	2	43.8	0.39*
32	2	2	25.0	0.00
33	3	0	18.8	0.38*
34	3	0	18.8	0.38*
35	5	0	31.3	0.63*

** Significance at 50.0% and 0.33 ** significance at 70.0%*

Examination of Table 2 reveals that:

i. 11 items out of 35 items have difficulty level equal to 50.0% and above. The 11 items are 1,2,3,8,12,13,22,23,24,25 and 27. Of these 11 items, only 4 of the items are multiple choice items.

ii. 6 items have difficulty level of 0% which imp1ies that no individual in both the upper and lower groups got the item correctly. This type of items should be revisited in terms of their suitability as they provide no information about the individual indifferences, Items 10, 14,16,20 28 and 29 according to Anastasi and Urbina (2007) provides no information to us about the testees.

iii. 12 items have discriminating index of 0.38 and above. Out of these 12 items, 8 are of the multiple choice items.

In the view of Ukwuije (2003: 108) test constructors should usual1y consider items with discriminating power = 0.4 and above as satisfactorily. In this study DP = 0.38 was considered satisfactorily as it is very near DP of 0.4.

iv, 8 items have DP equals to zero. This means that the percentage of both groups passing the item are equals. These items are 6,10,14, 15,16,20,28,29 and 32. Of these items, 5 are multiple choice items.

According to Ukwuije (2003: 108) on ideal difficulty level for an objective test item should not be less than 70%. If you go by this, only 4 items are accepted in terms of item difficulty level and they are items 1,12,13 and 23.

As Guilford (1984) has pointed out, the difficulty of an item varies for individuals. We do not have accurate information concerning an item's difficulty for an individual. All we know is that if he passes it, the item is less difficult than his ability to cope with it and if he fails it is more difficult than his ability to cope with it.

Evaluating the Effectiveness of Distracters

How well each distracter is operating can be determined by inspection and there is no need to calculate an index of effectiveness. In general, a good distracter is one that attracts more testees from the lower group than the upper group. Table 3: Response Analysis of Individual Items

		Response Options					
Items	**Group**	**A**	**B**	**C**	**D**	**E**	**Omitted**
4	Upper	1	0	1	2	1*	3
	Lower	4	1	0	I	0	2
5	Upper	0	0	4*	1	2	I
	Lower	1	0	4	I	1	1
6	Upper	1	1	2	1	2*	0
	Lower	2	1	1	2	2	0
14	Upper	8*	0	0	0	0	0
	Lower	8	0	0	0	0	0
15	Upper	2	0	2	2	1*	1
	Lower	1	0	1	3	2	1

22	Upper	0	0	0	0	8*	0
	Lower	4	0	0	0	3	1
23	Upper	8*	0	0	0	0	0
	Lower	4	1	0	1	0	2
24	Upper Lower	0 0	1 3	7 1	0 0	0 1	0 3
26	Upper	2	1	3	1	1	0
	Lower	1	0	2	0	1	4
27	Upper	0	8*	0	0	0	0
	Lower	0	1	0	1	I	5
28	Upper	4	0*	0	1	2	1
	Lower	1	0	1	0	1	5
29	Upper	1*	2	1	1	1	2
	Lower	0	0	1	3	I	-3
31	Upper	0	0	0	2	5*	I
	Lower	0	0	1	I	2	4
32	Upper	0	2	0	2	2*	2
	Lower	1	0	0	0	3	4
33	Upper	2	1	0	0	3*	2
	Lower	0	1	0	2	0	5
34	Upper	0	3*	1	1	2	1
	Lower	0	0	1	2	2	3
35	Upper	0	1	1	0	5*	1
	Lower	1	0	1	2	0	4

The * option is the correct answer to the item. The examination of table 3 reveals that

- In item 4, the correct key is E but the number of testees who omitted the items are more and those who choose Distractor are more than those who choose distractors C. Distracters in the item are not well written hence they are ineffective and should be reviewed.
- In item 5,both the upper and lower groups who choose the correct option are the same which means the distractor did not discriminate adequately between the upper and lower groups.
- Item 6 and 14 are just like item 5, in the sense that those who picked the correct option are the same.
- From both groups items 22, 23, 24,27,31,33,34 and 35 are very good items as the correct stem discriminate properly between the upper and lower groups. Each of the distractors in the item are good distracters because the lower group choose them more than the upper group.

Limitations of the Item Analysis of TNO

The item analysis of Test of Number Operations was faced with some limitations. They are as

follows:

- Some of the items in the Test are not multiple choice and a few of the items have more than one acceptable answer.
- The number of students used is quite few and the small size of the sample (30) used for analysis win falsify our results and lead to taking wrong conclusions.
- The examination of the effectiveness of the distracters is by inspection and this requires expertise in order to be able to make a meaningful evaluation of the distracters.

CONCLUSION AND RECOMMENDATION

The item analysis of the Test of Number of Operation was carried out and we observed the range of the actual score of the 30 students to be 12 (18-6) while the range of the true score is 14.4 (14.6-0.2). The difficulty level of all items was 50% and above and the rest 25 items were below 50%.

12 items of the test discriminate properly between the upper and lower groups while the other 23 items do not and need to be reviewed.

8 items of the test have distracters that distract effectively out of the 17 items that were analyzed. In conclusion the items in the Test of Number Operations should be worked upon to increase it, item statistics.

The result of this analysis should be compared with other analysis in different parts of the state and the sample should be increased.

REFERENCES

Aiken, L.R. (1979) *Psychological Testing and Assessment* (3rd Edition) Boston: Allyn and Bacon Inc. pp.28-49.

Anastasi A and Urbina, S. (2007) *Psychological Testing* (Seventh Edition) New Delhi: Pearson Education Inc. pp. 187-216

Badmus, G.A. (1977) Problems and issues on Criterion Referenced Measures: A cautionary Note to Curriculum Evaluator, *West African Journal of Education and Vocational Measurement* 3(3) 5-13

Badmus, G.A. and Omoifo, C.N. (1998). *Essential of Measurement and Evaluation in Education.* Benin City. Osasu Publishers pp92-95 Cronbach, Lee, J. (I *990).Essential of Psychological Testing* (fifth Edition). New York Harper Collins Publishers pp 177 - 178

Egbule J.F. (2007). *Principles of Psychological Testing* (Revised Edition) Lagos Functional Publishing Co.

Gronlund, N.E. (1976) *Measurement and Evaluation in Teaching* (3rd Edition New York. Macmillan Publishing Co. Inc.

Guilford, J.P. (1984) *Psychometric Methods.* New York: McGraw-Hill Inc. pp414~465.

Nunnal1y, Jum C. (1981) *Psychometric Theory.* New Delhi. Tata McGrawHill Publishing Co. Ltd pp256-265.

Okpala, P.N., Onocha, C.O. and Oyedeji O.A. (1993) *Measurement and Evaluation in Education.* Benin city Stirling - Horden Publishers (Nig. Ltd. pp73-76.

Ukwuije, R.P.I. (2003). *Test and Measurement for Teachers.* Port Harcourt: CELWIL Publishers pp 107 – 110.