International Journal of New Technology and Research (IJNTR)
ISSN: 2454-4116, Volume-5, Issue-12, December 2019 Pages 01-08

Item Analysis: An Evaluation of Multiple Choice Questions Based on Research Methodology in the Internal Examination

Lok Raj Sharma

***Abstract*— Item analysis is the process of collecting, summarizing, and using information from students' responses to assess the quality of multiple-choice questions (MCQs). The prime purpose of this research article is to analyze the twenty multiple choice question items regarding research methodology having asked to the Bachelor of Education (B. Ed) fourth year students in their internal examination conducted at Makawanpur Multiple Campus, Nepal in 2019. The researcher employed 27% students from the group of high achievers from the top and 27% students from the group of low achievers from the bottom by including 28 students from the population of 50 students. Each item was analyzed for difficulty index (DIF I), discrimination index (DI), and distracter effectiveness (DE). Difficulty index of 20 (100%) items was in the acceptable range (DIF I= 0.30–0.80). Discrimination index of 5 (25%) items was acceptable (DI= 0.20-0.29), that of 5 (25%) items was good (DI= 0.30-0.39) and that of 10 (50%) items was excellent (DI ≥0.40). Total 20 items had 60 distracters. 55 (91.67%) distracters were functional and 5(8.33%) distracters were non-functional. It was found that most of the multiple choice question items were reliable and valid. This article is useful to those who are involved in the field of teaching learning as well as question setting activities that concern the quality of the MCQS prepared for the examinees.**

***Index Terms*— Difficulty index, discrimination index, distractor effectiveness, item analysis, multiple choice questions.**

I. INTRODUCTION

One of the ways of evaluating the students' understanding of subject matters is through an assessment that is an essential phenomenon in the course of teaching and learning activities. Popham (2002) and Trice (2000) state that students' assessment and evaluation are an integral part of the teaching and learning process Assessing the students' understanding can be done through test items. Multiple choice questions are very common objective tests used in the faculty of education in Nepal. Such questions are also implemented in entrance examinations of medical sector, engineering, science, civil services and so on.

We can evaluate the students' understanding of subject matters and quality of multiple choice items through item analysis. Multiple choice questions (MCQs) are frequently used to assess students in different educational streams for objectivity and wide reach of coverage in a short period of time. This Study was done at Makawanpur Multiple Campus, Hetauda, Nepal in 2019. Total 20 MCQs with 4 alternative choices were prepared and 60 distracters of an internal examination of Research Methodology in English Education were analyzed by involving fourth year major English students.

The article writer adopted the general standard of taking 27% students from the group of high achievers from the top and 27% students from the group of low achievers from the bottom. The sample consists of 28 students from the population of 50 students.

A. *Objectives of the Study*

The key objective of the research study is:
To analyze the multiple choice question items asked to the fourth year students of Bachelor of Education in their internal examination of Research Methodology.

B. *Delimitations of the Study*

The research study was carried out under the following delimitations:

i. The study included only 28 students as the respondents.
ii. Only B.Ed. fourth year students were involved in the study.
iii. The study was delimited to only one campus.
iv. The questionnaire consisted of objective questions with only four alternatives in each question.

II. LITERATURE REVIEW

Literature review involves the definition of item analysis, elements of multiple choice question, concept and calculation of difficulty index, discrimination index and distracters.

A. *Item Analysis*

Item analysis is a statistical technique which is used for selecting and rejecting the items of the test on the basis of their difficulty index and discrimination index. Item Analysis is an important tool to increase the effectiveness of the test.

Item is a statement in the form of a question. Item analysis is one of the most important aspects of test construction. Item analysis is a general term for a set of methods used to evaluate test items. Items can be analyzed qualitatively in terms of their content and form and quantitatively in terms of their statistical properties.

Item analysis assesses the assessment tool for the benefit of both student and teacher. Item analysis is a process which examines student's responses to individual test items to assess the quality of these items and quality of test as a whole.

Lok Raj Sharma, Associate Professor of English Makawanpur Multiple Campus, Hetauda, Nepal

Gronlund (1993) opines that item analysis allows us to observe the item characteristics, and to improve the quality of the test. Lange and Mehrens (1967) assume that item revision allows us to identify items too difficult or too easy, items not able to differentiate between students who have learned the content and those who have not, or questions that have distracters not plausible. Singh, Gupta and Singh (2009) define item analysis as a process which assesses the quality of those items and of the test as a whole. Zubairi and Kassim (2006) and Sim and Rasiah (2006) assert that item analysis provides feedback to teachers for necessary modifications in MCQs to make it suitable for the exam. While some MCQs are edited, some are deleted based on the analysis.

Item analysis is a post validation procedure that characterizes every MCQ and its Distracters by assigning a numerical value to it in the form of a difficulty index, a discrimination index and distractor efficiency. Sharma (2000) and Freeman (1962) consider that the quality of a test depends upon each items of a test.

B. *Elements of Multiple Choice Questions*

A MCQ is composed of a stem and several options. The question or the statement in the sense of a question is called the stem. The correct option is called the key while the incorrect alternatives are called the distracters. MCQs are used mostly for comprehensive assessment at the end of academic sessions and they provide feedback to the teachers on their educational actions. Designing MCQ is a complex and time-consuming process in a multidisciplinary, integrated curriculum. MCQ needs to be tested for the standard or quality. Item analysis examines the student responses to individual test items (MCQs) to assess the quality of those items and test as a whole.

Brown (2004) considers that multiple-choice items are described as receptive or selective. Öztürk (2007) highlights the importance of multiple choice items and remarks that multiple-choice items seem to be reliable compared with other types of tests which are negatively affected by subjectivity. Multiple choice-items were significantly easier and less discriminating than free response tasks. Hughes (2003) emphasizes some weaknesses of multiple-choice items stating that this technique only tests recognition knowledge which is a lower mental skill according to Bloom's taxonomy (1956).

C. *Difficulty Index (DIF I)*

The difficulty index (DIF I) of an item is defined as proportion or percentage of the examinees who correctly answered a given test item. It is one of the key parameters of item analysis. It ranges from 0 to 1 or (0% to 100%). We can also use the following formula to calculate the difficulty Index.

$$\text{DIF I} = \frac{RH+RL}{NH+NL} \quad \text{or} \quad = \frac{RH+RL}{N}$$

Where, RH→The number of right answers in the higher group; RL→The number of right answers in the lower group; NH→The number of examinees in the higher group; N→Total number of examinees and NL→ The number of examinees in the lower group.

In case of non-response examinees, we should use the following formula for determining DIF I.

$$\text{DIF I} = \frac{RH+RL}{NH+NL-NR}$$

Where, NR→ The number of non-response examinees.

D. *Discrimination Index (DI)*

Index of discrimination is that ability of an item on the basis of which the discrimination is made between the superior (the group of high achievers) and the inferior (the group of low achievers). It ranges from -1 to +1. Gajjar, Sharma, Kumar and Rana (2014) define item discrimination as "the ability of an item to differentiate between students of higher and lower abilities" (p.18).

We can use the following Formula for determining the discrimination Index:

$$\text{DI} = \frac{RH-RL}{NH} \quad \text{or, DI} = \frac{RH-RL}{NL} \quad \text{or,} \quad \text{DI} = \frac{2\,(RH-RL)}{N}$$

E. *Distracter*

A distracter is a wrong option in the multiple choice question. It is designed to see whether the person being tested can notice the difference in a test or not. A distracter is considered to be a good distracter when it attracts more examinees from the group of low achievers than the group of high achievers. According to Malau-Aduli and Zimitat (2012), a distractor that fails to "attract any examinees is dysfunctional, does not assist in the measuring of educational outcomes, adds nothing to the item or the test (psychometrically) and has negative impact upon learners" (p.927).

There are two types of distracter. They are: non-functional distracter (NFD) and functional distracter (FD). Non-functional distracter (NFD) in an item is the option, other than the correct option selected by less than 5% of students and the functional or effective distracter is the option selected by 5% or more. It may not exactly be of 0 frequency choice. The percent of a distracter can be calculated by using the following formula:

$$\textbf{Percent of a distracter} = \frac{\text{Number of examinees of the Distracter}}{\text{Total Number of examinees}} \times 100\ \%$$

F. *Effectiveness of Distracters / Distracter Efficiency (DE)*

Distracter efficiency (DE) for any item ranges from 0 to 100% and is determined on the basis of the number of NFDs in an item. It means DE is expressed as 0%, 33.3%, 66.6% and 100% depending on the number of NFDs.

Number of Non-functional Distracters	Distracter Efficiency (DE)
3 NFDs	0.00 %
2NFDs	33.33 %
1NFD	66.66 %
0 NFD	100 %

III. METHODOLOGY

Methodology involves research design, population, sampling design, sample size, variables, nature, validity of instrument, nature and sources of data and analysis and interpretation of data.

A. Research Design

A cross-sectional survey design was used to carry out the study. The researcher collected the primary data from the questionnaire to investigate test items asked to the Bachelor of Education students in their internal research examination.

B. Population / Universe

The population of the study consisted of 50 Bachelor of Education students studying at Makawanpur Multiple Campus, Nepal in the Academic Year 2019.

C. Sampling Design

The researcher selected 28 students by following the general standard of selecting 27% respondents from the group of high achievers from the top and 27% respondents from the group of low achievers from the bottom after maintaining the scores in descending order.

D. Sample Size by Gender

The sample size of the research study consisted of 12 boys and 16 girls studying at the campus.

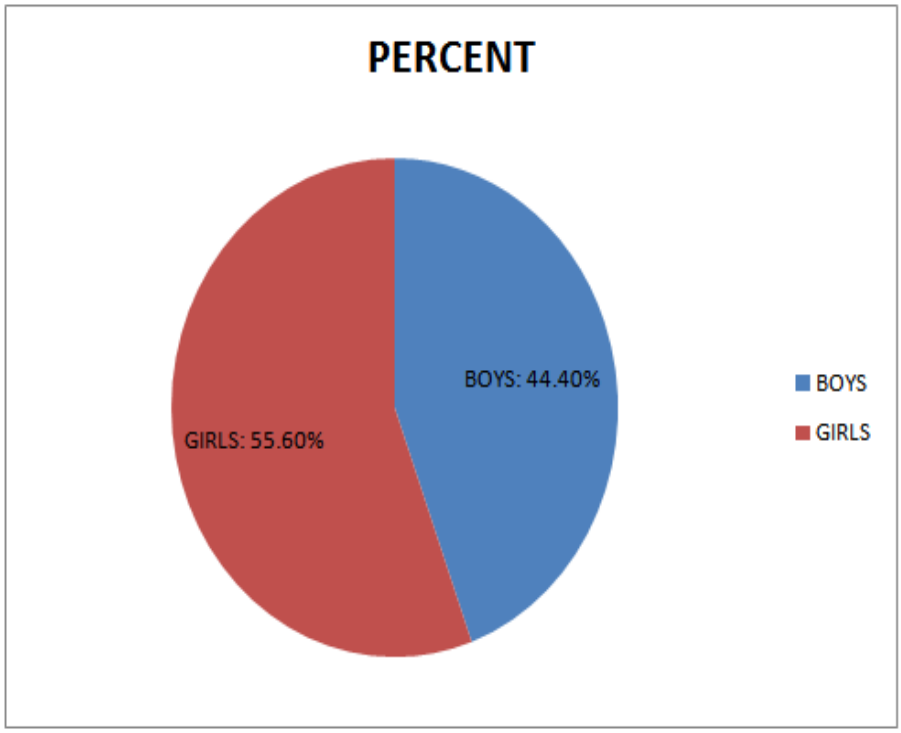

Figure 1: Percent of Students by Gender

This figure shows that the percent of boys was a bit smaller than that of the girls.

E. Variables in the Study

The researcher took twenty multiple choice items with four alternatives (1 correct answer and 3 distracters in each and every question) as major variables.

F. Validity and Reliability of the Instruments

The validity of questionnaire was conducted by using Pearson Product moment correlations in SPSS, where the score of each item was correlated with the total score. The Pearson Product moment correlation count value was greater than the corresponding Pearson correlation critical value ***/r/>rc= 0.374*** of all questions except question numbers 18 and 20. It indicates that most of the questions were valid. The Split-half reliability test was used to check the reliability of the questions. The overall internal consistency of questions was 0.954 which indicates that the questions were highly reliable.

G. Nature and Source of Data

The researcher exploited ratio data to carry out the research study. Students' performance in research methodology was tested through the use of objective questions regarding their understanding of the research methodology. The primary source of data was questionnaire. The secondary source of data included books, journal articles, web-sites etc.

IV. ANALYSIS AND INTERPRETATION OF DATA

All the data were analyzed by using Statistical Package for Social Sciences (SPSS) 20 version. The researcher used a series of twenty multiple choice questions asked in the internal examination to the Bachelor of Education fourth year students to analyze the test items. The researcher used the frequency and percent statistics to analyze the test items by focusing the difficulty Index, the discrimination index and the effectiveness of distracters in each test item.

Table 1: Number of examinees who answered rightly in both groups

Items	Right Answers from the Group of High Achievers (RH)	Right Answers from the Group of Low Achievers (RL)	Total Right Answers
Item No. 1	10	4	14
Item No. 2	8	2	10
Item No. 3	9	2	11
Item No. 4	10	6	16
Item No. 5	10	5	15
Item No. 6	10	4	14
Item No. 7	9	2	11
Item No. 8	10	2	12
Item No. 9	10	6	16
Item No. 10	10	5	15
Item No. 11	11	4	15
Item No. 12	11	2	13
Item No. 13	9	2	11
Item No. 14	11	6	17
Item No. 15	10	5	15
Item No. 16	10	5	15
Item No. 17	10	4	14
Item No. 18	10	6	16
Item No. 19	9	5	14
Item No. 20	12	8	20

A. Calculation of Difficulty Index and Discrimination Index

The difficulty index and the discrimination index of each question item can be computed by using the following formulae:

Difficulty Index (DIF I) = $\frac{RH+RL}{N}$ & Discrimination Index (DI) = $\frac{2\,(RH-RL)}{N}$

Table 2: Difficulty index, discrimination index and number of non-functional distracters

MCQ Item Number	Difficulty Index (DIF I)	Discrimination Index (DI)	Number of NFDs	Effectiveness of Distracters (EDs)
Item No.1	0.50	0.42	0	100.00 %
Item No.2	0.35	0.42	0	100.00 %
Item No.3	0.39	0.50	1	66.66 %
Item No.4	0.57	0.28	0	100.00 %
Item No.5	0.53	0.35	0	100.00 %
Item No.6	0.50	0.42	0	100.00 %
Item No.7	0.39	0.50	0	100.00 %
Item No.8	0.42	0.57	1	66.66 %
Item No.9	0.57	0.28	0	100.00 %
Item No.10	0.53	0.35	0	100.00 %
Item No.11	0.53	0.50	0	100.00 %
Item No.12	0.46	0.64	0	100.00 %
Item No.13	0.39	0.50	1	66.66 %
Item No.14	0.60	0.35	0	100.00 %
Item No.15	0.53	0.35	0	100.00 %
Item No.16	0.53	0.35	0	100.00 %
Item No.17	0.50	0.42	0	100.00 %
Item No.18	0.57	0.28	1	66.66 %
Item No.19	0.50	0.28	0	100.00 %
Item No.20	0.71	0.28	1	66.66 %

B. Interpretation of Difficulty Index

The researcher has employed the following table to analyze the difficulty index of each question item.

Table 3: Evaluation of difficulty index

S. N.	Difficulty Index	Item Evaluation	Recommendation
1.	< 0.30	Most difficult	Remove
2.	0.30-0.49	Difficult	Keep
3.	0.50-0.69	Moderately difficult	Keep
4.	0.70-0.80	Moderately Easy	Keep
5.	> 0.80	Easiest	Remove

6 question items fell in the range of the difficulty index of (0.30 – 0.49). It shows that 6 questions were difficult. 13 question items fell in the range of the difficulty index of (0.50 – 0.69). It indicates that 13 questions were moderately difficult. 1 question item fell in the range of the difficulty index of (0.70 – 0.80). This question was moderately easy. The examiner was recommended to maintain question items in the list of questionnaire.

C. Interpretation of Discrimination Index

The researcher used the following table to analyze and interpret the discrimination index of each question item.

Table 4: Evaluation of discrimination index

S. N.	Discrimination Index	Numbers of Items	Percent	Evaluation of Item	Action
1.	**Negative**	**0**	**0**	**Worst/ Defective**	**Definitely Discard**
2.	**< 0.20**	**0**	**0**	**Poor**	**Revise / Discard**
3.	**0.20- 0.29**	**5**	**25%**	**Acceptable**	**Keep**
4.	**.30- 0.39**	**5**	**25%**	**Good**	**Keep**
5.	**≥ 0.40**	**10**	**50%**	**Excellent**	**keep**

This table demonstrates that 5 (25%) question items that lay in the discrimination index of (0.20-0.29) were acceptable for the inclusion in the list, 5 (25%) question items that lay in the discrimination index of 0.30-0.39 were good items and 10 (50%) question items were excellent items that must have been kept in the list.

D. Calculation of the Effectiveness of Distracters / Distracter Efficiency in Each Item

Distracter efficiency is determined on the basis of the number of NFDs in an item. It means DE is expressed as 0%, 33.3%, 66.6% and 100% depending on the number of NFDs. The percent of each distracter was calculated by using the following formula:

$$\textbf{Percent of a distracter} = \frac{\text{Number of examinees of the Distracter}}{\text{Total Number of examinees}} \times 100\ \%$$

Table 5: Distracter efficiency in each item

Number of Non-functional Distracters	Question No.	Distracter Efficiency (DE)
1NFD	3,8,13,18,20	66.66 %
0 NFD	1,2,4,5,6,7,9,10,11,12,14,15,16,17,19	100 %

There were 60 distracters in the question items. 55 (91.67%) distracters were functional and 5(8.33%) distracters were non-functional. It means most of the distracters were effective in test items. It can be said that the effectiveness of distracters in 5 questions was 66.66% and that of in 15 questions was 100.00%.

E. Calculation of Mean Scores of the Difficulty Index and the Discrimination Index

The researcher computed the mean scores of the difficulty index and the discrimination index to evaluate the items as a whole.

Table 6: Mean and standard deviation of difficulty index

Aspect	No. of Items	Range	Minimum	maximum	Mean	Standard Deviation
Difficulty Index	20	.35	.35	.70	**0.50**	.08

The mean score of the difficulty index was 0.50 that lay between 0.50- 0.69. It shows that the test items were moderately difficult and they should be kept in the list of test items.

Table 7: Mean and standard deviation of discrimination index

Aspect	No. of Items	Range	Minimum	maximum	Mean	Standard Deviation
Discrimination Index	20	.36	.28	.64	**0.40**	.10

This table shows that the mean score of the discrimination index was 0.40. It reveals that all the multiple choice questions were excellent from the perspective of the standard of discrimination index value. It suggests that the examiner should keep these items in the list of the test.

V. DISCUSSIN AND RESULT

All the multiple choice questions fell in the difficulty index range between 0.30-0.80 and that was the range of accepting the test items. 6 question items fell in the range of the difficulty index of (0.30 – 0.49). It shows that 6 questions were difficult. 13 question items fell in the range of the difficulty index of (0.50 – 0.69). It indicates that 13 questions were moderately difficult. 1 question item fell in the range of the difficulty index of (0.70 – 0.80). This question was moderately easy. The mean score of the difficulty indices of the test items was 0.50. Similarly, 5 (25%) question items that lay in the discrimination index of 0.20-0.29 were acceptable for the inclusion in the list, 5 (25%) question items that lay in the discrimination index of 0.30-0.39 were good items and 10 (50%) question items were excellent items that must have been kept in the list. The mean score of the discrimination indices of the question items was 0.40 and the scale shows that question items were excellent. 55 (91.67%) distracters were functional, whereas 5 (8.33%) distracters were non-functional.

VI. CONCLUSION

Item analysis is a prominent procedure performed after the examination for providing information regarding the reliability and validity of an item/test by calculating difficulty index, discrimination index, distracter efficiency, and their interrelationship. An ideal item (MCQ) will be the one which has average difficulty index between 0.31 and 0.60, high discrimination (*DI* > 0.20), and maximum distractor efficiency (100%) with three functional distracters. Items analyzed in the study were neither too easy nor too difficult (mean difficulty index = 0.50), and overall discrimination index was 0.4, which was excellent.

In this study, the majority of items fulfilled the criteria of acceptable difficulty and good discrimination which means the MCQs selected were of good quality. Moderately easy/difficult had maximum discrimination ability. Very easy and very difficult items displayed poor discrimination. Such items were recommended to be reviewed and reconstructed.

The results of this study should initiate a change in the way MCQ test items are selected for any examination, and there should be proper assessment strategy as a part of the curriculum development. The researcher likes to recommend the other researchers to carry out further item analysis after each examination to identify the areas of potential weakness in the formation of MCQ items to improve the standard of assessment.

VII. ACKNOWLEDGEMENTS

I would like to thank the Bachelor of Education fourth year students of Makawanpur Multiple Campus for their active participation in responding to the questionnaire so that I could collect the data required for my study.

REFERENCES

[1] Brown, H. D. (2004). *Language assessment: Principles and classroom practices.* White Plains, NY: Pearson Edu¬cation.
[2] Freeman, F. (1962). *Theory and practice of psychological testing.* New Delhi: Oxford & Ibh publishing.
[3] 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 stu¬dents of Ahmedabad, Gujarat . *Indian Journal of Com¬munity Medicine, 39*(1), 17-22.
[4] Gronlund, N. E. (1998). *Assessment of student achievement. 6th edition.* Boston: Allyn and Bacon.
[5] Hughes, A. (2003). *Testing for language teachers* (2 nd ed.). Cambridge: Cambridge University Press.
[6] Lange, A., Lehmann, I.J. & Mehrens, W.A. (1967). Using item analysis to improve tests. *Journal of Educational Measurement, 4*(2), 65-68. Retrieved from http://www.jstor.org/stable/1434299.
[7] Malau-Aduli, B. S., & Zimitat, C. (2012). Peer review im¬proves the quality of MCQ examinations . *Assessment & Evaluation in Higher Education, 37*(8), 919-931.
[8] Öztürk, M. (2007). Multiple-choice test items of foreign language vocabulary. *Eğitim Fakültesi Dergisi, 20*(2), 399-426.
[9] Popham, W. J. (2008). *Classroom assessment: What teachers need to Know (5th ed.).* Boston: Allyn and Bacon.
[10] Sharma, S. (2008). *Modern teaching strategies. New Delhi: Omsons Publications.* New Delhi: Omsons Publications.
[11] Sim S.M.& Rasiah R. I. (2006). Relationship between item difficulty and discrimination indices in true/falsetype multiple choice questions of a para-clinical multidisciplinary paper. *Annals Academy of Medicine Singapore, 35*, 67-71.
[12] Singh T, Gupta P, Singh D. (2009). Test and item analysis . *Principles of Medical Education 3rd ed*, 70-77.
[13] Trice, A. D. (2000). *A handbook of classroom assessment.* New York: Longman.
[14] Zubairi A.M. & Kassim N.L. (2006). Classical and Rasch analysis of dichotomously scored reading comprehension test items. *Malaysain Journal of English Language Teaching Research, 2*, 1-20.

Appendix- I
Objective Questions Based on Research Methodology

Q.1 In an experimental design, the dependent variable is:
a. The one that is not manipulated and in which any changes are observed.
b. The one that is manipulated in order to observe any effects on the other.
c. A measure of the extent to which personal values affect research.
d. An ambiguous concept whose meaning depends on how it is defined.

Q.2 What is a cross-sectional design?
a. A study of one particular section of society, e.g. the middle classes.
b. One that is devised when the researcher is in a bad mood.
c. The collection of data from more than one case at one moment in time.
d. A comparison of two or more variables over a long period of time.

Q.3 Research that uses qualitative methods for one phase and quantitative methods for the next phase is known as:
a. Action research
b. Mixed-method research
c. Quantitative research
d. Pragmatic research

Q.4 Which research paradigm is most concerned about generalizing its findings?
a. Quantitative research
b. Qualitative research
c. Mixed-methods research
d. Descriptive research

Q.5 A variable that is presumed to cause a change in another variable is called:
a. An intervening variable
b. A dependent variable
c. An independent variable
d. A numerical variable

Q.6 Which of the following can be described as a nominal variable?
a. Annual income
b. Age
c. Annual sales
d. Geographical location of a firm

Q.7 A positive correlation occurs when:
a. Two variables remain constant
b. Two variables move in the same direction
c. One variable goes up and the other goes down
d. Two variables move in opposite directions

Q.8 Qualitative research is used in all the following circumstances, EXCEPT:
a. It is based on a collection of non-numerical data such as words and pictures
b. It often uses small samples
c. It uses the inductive method
d. It is typically used when a great deal is already known about the topic of interest

Q.9 In an experiment, the group that does not receive the intervention is called:
a. The experimental group
b. The participant group
c. The control group
d. The treatment group

Q.10 What do we call data that are used for a new study but which were collected by an earlier researcher for a different set of research questions?
a. Secondary data
b. Field notes
c. Qualitative data
d. Primary data

Q.11 When each member of a population has an equal chance of being selected, this is called:
a. A snowball sample
b. A stratified sample
c. A random probability sample
d. A non-random sample

Q.12 Which one of the following statistical tools is a measure of dispersion?
a. Mean
b. Median
c. Standard deviation
d. Frequency

Q.13 When designing a questionnaire it is important to do each of the following EXCEPT **a. Piloting the questionnaire** **b. Avoiding jargon** **c. Avoiding double questions** **d. Using leading questions**
Q.14 An ordinal scale is: **a. The simplest form of measurement** **b. A scale with an absolute zero point** **c. A rank-order scale of measurement** **d. A scale with equal intervals between ranks**
Q.15 The 'reliability 'of a measure refers to the researcher asking: **a. Does it give consistent results?** **b. Does it measure what it is supposed to measure?** **c. Can the results be generalized?** **d. Does it have face reliability?**
Q.16 Interview questions should: **a. Lead the respondent** **b. Probe sensitive issues** **c. Be delivered in a neutral tone** **d. Test the respondents' powers of memory**
Q. 17 A Type 1 error occurs in a situation where: **a. The null hypothesis is accepted when it is in fact true** **b. The null hypothesis is rejected when it is in fact false** **c. The null hypothesis is rejected when it is in fact true** **d. The null hypothesis is accepted when it is in fact false**
Q.18 Mean, Median and Mode are : **a. Measures of deviation** **b. Ways of sampling** **c. Measures of central tendency** **d. Research designs** **a.**
Q. 19 Action research means **a. A longitudinal research** **b. An applied research** **c. A research initiated to solve an immediate problem** **d. A research with socioeconomic objective**
Q.20 which of the following is non-probability sampling? **a. Snowball** **b. Random** **c. Cluster** **d. Stratified**