THE UNITED REPUBLIC OF TANZANIA

NATIONAL EXAMINATION COUNCIL OF TANZANIA

DIPLOMA IN SECONDARY EDUCATION EXAMINATION

762 EDUCATIONAL RESEARCH, MEASUREMENT AND EVALUATION

Time: 3 Hours. **ANSWER** **Year: 2014 a.m.**

Instructions

1. This paper consists of sections A and B with a total of **sixteen (16)** questions
2. Answer **all** questions in sections A and **four (4)** questions from section B.
3. Question **11** is **compulsory**.
4. Section A carries **forty (40)** marks and section B carries **sixty (60)** marks.
5. Cellular phones and unauthorized materials are **not allowed** in the examination room.
6. Non-Programmable calculators may be used.
7. Write your **Examination Number** on every page of your answer booklet(s).

1. Distinguish the following terms:
(a) Predictive validity and face validity.

Predictive validity refers to the extent to which a test score can forecast or predict future performance on a related task. For example, university entrance exams should predict how well a student will perform in college. It involves a statistical relationship between test scores and future behavior or performance.

Face validity, on the other hand, refers to the extent to which a test appears to measure what it claims to measure, based on subjective judgment. It is not statistically tested but depends on the impression of test-takers or experts. A test has face validity if it looks appropriate and relevant on the surface.

(b) Ordinal scale and interval scale.

Ordinal scale is a level of measurement where items are ranked in a specific order, but the intervals between the ranks are not necessarily equal. For example, positions in a race (1st, 2nd, 3rd) indicate order but not the actual time differences between runners.

Interval scale not only ranks items but also provides equal spacing between values. However, it lacks a true zero point. For example, temperature measured in Celsius is an interval scale, where the difference between 20°C and 30°C is equal to the difference between 30°C and 40°C, but 0°C does not mean 'no temperature'.

2. (a) What is the difference between an inductive theory and a deductive theory?

Inductive theory is developed by observing specific data or patterns and then forming general principles or theories from them. It moves from particular observations to broader generalizations. For example, a researcher may observe that students who sleep well perform better and then form a theory about sleep and academic performance.

Deductive theory begins with a general theory or principle and then tests it through specific observations or experiments. It moves from the general to the specific. For example, a theory that says motivation improves performance can be tested by observing if motivated students actually score higher.

(b) State two characteristics of a useful theory.

A useful theory should be testable, meaning it can be examined through research and empirical data to prove or disprove its claims.

It should also be applicable, meaning it can be used to explain or solve real-world problems and guide further research or practice.

3. (a) State three characteristics of a nominal scale of measurement.

A nominal scale classifies data into distinct categories without any quantitative value. The categories are mutually exclusive and exhaustive.

There is no inherent order among the categories. For example, gender or religion categories have no ranking.

It uses names, labels, or symbols to represent different groups or classes, such as 1 = Male, 2 = Female.

(b) Why should a teacher choose the quartile deviation or standard deviation as the index of variability in order to minimize the effect of extreme score?

Quartile deviation focuses on the middle 50% of scores, ignoring the highest and lowest values. This makes it less affected by extreme scores and more representative of the central trend.

Standard deviation uses all data points but spreads the weight evenly, making it more stable and less influenced by individual outliers compared to simple range. It gives a more accurate reflection of variability in most data sets.

4. Outline four characteristics of matching-item type of questions.

Matching items consist of two columns, one with questions or prompts and another with answers—which students must pair correctly.

They are efficient for testing knowledge of relationships, definitions, or associations, especially when dealing with many related concepts.

They are objective in nature since answers are clearly right or wrong, reducing scorer bias.

They are time-efficient, allowing many items to be tested within a short time, useful for large coverage of content.

5. Suggest four criteria of a good educational research problem.

A good research problem should be clearly stated, so that its purpose and scope are understandable.

It must be researchable, meaning it can be investigated using available methods and data.

It should be significant, addressing real issues that contribute to knowledge or improve educational practice.

It must be feasible, considering the time, resources, and skills available to the researcher.

6. Briefly explain four importance of literature review in research proposal.

It helps identify gaps in existing knowledge, allowing the researcher to contribute new insights rather than repeating previous work.

It guides the development of research questions and methodology by reviewing what has worked in previous studies.

It provides a theoretical foundation and context, linking current research with established theories.

It helps avoid duplication and refine the research focus by understanding what has already been studied.

7. Briefly explain four characteristics of instructional objectives.

Instructional objectives must be specific, clearly stating what the learner is expected to do.

They must be measurable, using observable behaviors that can be assessed reliably.

They should be achievable, aligned with learners' abilities and resources available.

They must be relevant, reflecting curriculum goals and the intended learning outcomes.

8. (a) Distinguish the terms assessment and evaluation.

Assessment refers to the process of collecting information about student learning through tests, assignments, or observations. Its main purpose is to provide feedback and improve learning.

Evaluation is broader and involves making judgments about the quality or value of a program, lesson, or performance. It includes assessment results but also considers goals, effectiveness, and outcomes for decision-making.

(b) Name four ways of estimating test reliability.

One way is test-retest reliability, which involves administering the same test twice to the same group and comparing the results.

Another is split-half reliability, where the test is divided into two parts, and the consistency of results between the halves is measured.

Internal consistency, often measured by Cronbach's alpha, assesses how closely related the items in a test are.

Inter-rater reliability compares the consistency of scores given by different evaluators to the same responses.

9. (a) Why should hypothesis be clearly stated before research is initiated?

Stating the hypothesis clearly provides a focused direction for the study and helps in selecting appropriate methods and tools for data collection.

It allows the researcher to make specific predictions and design the study in a way that tests the relationship between variables.

It also helps in the interpretation of results, as the hypothesis serves as a basis for comparing findings and drawing conclusions.

(b) Give the meaning of research design.

A research design is the overall plan or structure of a research study. It outlines how data will be collected, analyzed, and interpreted to answer research questions or test hypotheses.

(c) Educational research can be categorized into two major designs. Differentiate the two categories of research designs based on their functional definitions.

Quantitative research design involves collecting numerical data and analyzing it statistically to test hypotheses or measure variables. It focuses on objectivity and measurable outcomes.

Qualitative research design uses non-numerical data like interviews or observations to explore phenomena in depth. It focuses on understanding meanings, experiences, and contexts from participants' perspectives.

10. (a) Define the term reliability.

Reliability refers to the consistency or stability of measurement results. A test is considered reliable if it produces the same results under consistent conditions over time or across different raters.

(b) Briefly describe three methods used in estimating reliability of a test.

Test-retest method involves giving the same test to the same group twice and correlating the scores to measure consistency over time.

Split-half method involves dividing a test into two equal parts and correlating the scores on each half to determine internal consistency.

Alternate form method requires creating two equivalent versions of a test and administering both to the same group to see if the scores remain consistent across forms.

11. Study carefully the scores of 15 students obtained from Foundations of Education test and answer the questions that follow: 45, 58, 49, 60, 36, 50, 60, 75, 45, 45, 21, 54, 70, 60 and 87.
(a) Find:
(i) Mode
(ii) Median
(iii) Range
(iv) Mean (round off the answer to the nearest whole number).

(i) Mode: The most frequently occurring score is 45 and 60, both appearing 3 times each, hence it is bimodal.

(ii) Median: Arranging the scores in ascending order: 21, 36, 45, 45, 45, 49, 50, 54, 58, 60, 60, 60, 70, 75, 87. The middle value (8th position) is 54, so the median is 54.

(iii) Range: Highest score 87 minus lowest score 21 = 66.

(iv) Mean:
Sum of all scores = 45 + 58 + 49 + 60 + 36 + 50 + 60 + 75 + 45 + 45 + 21 + 54 + 70 + 60 + 87 = 815.
Mean = 815 ÷ 15 = 54 (rounded to nearest whole number).

(b) Calculate the standard deviation (round off the answer to the nearest whole number).

Step 1: Find the mean = 54.
Step 2: Subtract the mean from each score, square the result:
$(45 - 54)^2 = 81$
$(58 - 54)^2 = 16$
$(49 - 54)^2 = 25$
$(60 - 54)^2 = 36$
$(36 - 54)^2 = 324$
$(50 - 54)^2 = 16$
$(60 - 54)^2 = 36$
$(75 - 54)^2 = 441$
$(45 - 54)^2 = 81$
$(45 - 54)^2 = 81$
$(21 - 54)^2 = 1089$
$(54 - 54)^2 = 0$
$(70 - 54)^2 = 256$
$(60 - 54)^2 = 36$
$(87 - 54)^2 = 1089$

Step 3: Sum of squared deviations = 3707.
Step 4: Variance = 3707 ÷ 15 = 247.13.
Step 5: Standard deviation = √247.13 = 16 (rounded).

(c) Write the formula of obtaining z-score and T-score.

Z-score formula: Z = (X − Mean) ÷ Standard Deviation.
T-score formula: T = (Z × 10) + 50.

12. Explain any five factors to be considered when constructing a matching items test.

The first factor is clarity of instructions. Students should clearly understand how to match items, whether answers can be used once or multiple times.

The second factor is logical relationship between columns. Items should be related by a common theme, such as terms and their definitions, to avoid confusion.

The third factor is length balance. The two columns should be of manageable length, avoiding an overwhelming number of choices.

The fourth factor is avoiding grammatical clues. The phrasing of items should not reveal the correct match.

The fifth factor is content relevance. The items should align with learning objectives and accurately reflect the subject matter taught.

13. Analyse five criteria that should be used in the process of evaluating the significance of a research problem.

The first criterion is relevance. The problem should address an important issue within the field of study, contributing to academic or practical knowledge.

The second is feasibility. It should be possible to research the problem given the available time, resources, and expertise.

The third is clarity. The problem should be stated in precise, specific terms that leave no ambiguity.

The fourth is originality. It should offer new insights or approaches rather than duplicating previous work.

The fifth is ethical soundness. The problem should be researchable without causing harm to participants or violating ethical standards.

14. Questionnaire is one of the best instruments for data collection. Examine six qualities of a good questionnaire for data collection in assessment.

A good questionnaire should be clear, using simple and understandable language to avoid misinterpretation.

It should be concise, keeping the number of questions manageable to maintain respondent engagement.

It must be relevant, with each question directly related to the research objectives.

It should be unbiased, avoiding leading questions that influence the respondent's answer.

It must be logically organized, grouping related questions together to maintain flow.

It should be pre-tested to identify and correct any issues before full administration.

15. (a) Explain six advantages of using interview as one of the assessment tools.

Interviews allow for in-depth data collection, capturing detailed responses that questionnaires may miss.

They enable clarification of questions, reducing misunderstandings.

They capture non-verbal cues such as tone and body language, adding depth to the data.

They are flexible, allowing the interviewer to adapt questions based on responses.

They build rapport, encouraging participants to share more openly.

They can explore sensitive topics more effectively than written methods.

(b) What are the two components of the basic teaching model?

The first component is the instructional objectives, which specify the desired learning outcomes.

The second is the learning activities, which are the strategies, materials, and methods used to help learners achieve the objectives.

16. (a) Give six factors to consider in constructing essay-type items.

The first factor is clarity of wording, ensuring students understand exactly what is required.

The second is defining the scope of the answer, specifying limits to avoid irrelevant details.

The third is aligning with instructional objectives, ensuring the item measures intended skills or knowledge.

The fourth is appropriateness to the learners' level, avoiding overly difficult or trivial topics.

The fifth is allocating adequate time for answering, ensuring the question can be completed within the test period.

The sixth is preparing a marking scheme to ensure fair and consistent grading.

(b) Explain three advantages and three disadvantages of short answer questions.

Advantages:
They are easy to construct compared to essays.
They reduce guessing compared to multiple choice questions.
They can cover a wide range of content in a short time.

Disadvantages:
They may encourage rote memorization.

Scoring can be subjective if answers are not standardized.
They provide limited scope for students to express depth of understanding.

17. (a) Explain in detail three functions of standardized test scores.

They allow comparison of student performance across different schools or regions, providing a common benchmark.

They help in placement decisions, such as identifying students for special programs or advanced classes.

They provide data for evaluating educational programs and curricula, helping policymakers make informed decisions.

(b) Explain four importance of evaluation in educational institutions.

Evaluation helps in measuring student learning outcomes to assess instructional effectiveness.

It identifies strengths and weaknesses in both teaching and learning processes.

It provides feedback to students, guiding their improvement.

It informs curriculum development and policy decisions for better education quality.

18. (a) Analyse five characteristics of the usable hypothesis.

A usable hypothesis must be clear, using precise terms that leave no room for confusion.

It should be testable, allowing verification through observation or experimentation.

It must be specific, focusing on a defined relationship between variables.

It should be consistent with existing knowledge or theory, making it logical and credible.

It must be feasible, meaning it can be investigated with available resources and within a reasonable timeframe.

(b) State one hypothesis based on each of the following research questions:
(i) Do teachers' expectations of children's intellectual performance have any effect on their actual performance?

Hypothesis: Teachers' high expectations positively influence children's intellectual performance.

(ii) Does preschool training reduce the educational gap separating advantaged and disadvantaged children before they enter first grade?

Hypothesis: Preschool training significantly reduces the educational gap between advantaged and disadvantaged children before first grade.