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: 2005 a.m.**

Instructions

1. This paper consists of sections A, B and C.
2. Answer **all** questions in sections A, **two (2)** questions from section B and **one (1)** question from section C.
3. Question **11** is **compulsory**.
4. Section A carries 36 marks, section B carries 40 marks and section C carries 24 marks
5. Cellular phones and unauthorized materials are **not allowed** in the examination room.
6. Write your **Examination Number** on every page of your answer booklet(s).

1. Define the term reliability as used in educational testing.

Reliability in educational testing refers to the consistency of a test in producing similar results when administered under the same conditions on different occasions.

It ensures that a test measures learning outcomes accurately without being affected by random errors or external factors.

A reliable test allows educators to make dependable decisions based on student performance.

It also gives confidence that differences in scores are due to differences in learners' abilities rather than flaws in the testing process.

2. Mention four characteristics of a good research instrument.

A good research instrument must be valid, meaning it measures exactly what it is intended to measure.

It should be reliable, producing consistent results when used repeatedly under the same conditions.

It must be clear and easy to use so that both the researcher and participants understand how to respond correctly.

It should be practical, meaning it can be used within the available time, budget, and resources.

3. State four purposes of measurement in education.

Measurement is used to determine the level of achievement of learners in relation to set objectives.

It provides data that helps teachers identify the strengths and weaknesses of students.

It is used for making decisions on promotion, placement, or certification of learners.

It helps evaluate the effectiveness of teaching methods and learning materials used in the classroom.

4. Give four reasons for establishing test norms.

Test norms allow the comparison of a student's score with the performance of a larger reference group.

They help interpret scores in a meaningful way by showing how a learner stands relative to peers.

They provide a basis for identifying students who may need additional support or enrichment.

They guide educational decisions by establishing performance standards for different groups of learners.

5. List three limitations of using observation as a data collection method.

Observation can be time-consuming, especially when trying to gather enough data for meaningful conclusions.

It may lead to observer bias, where the observer's personal opinions influence the interpretation of behavior.

Some behaviors may not occur during the observation period, meaning important data could be missed.

6. State four advantages of using standardized tests in schools.

Standardized tests allow for fair comparison between students from different schools or regions.

They provide reliable and valid measures of student performance because they are developed under strict procedures.

They are useful for making educational policy decisions based on comparable data.

They can identify trends in performance over time, helping in planning educational improvements.

7. Mention four steps in preparing a table of specifications.

The first step is identifying the learning objectives to be measured in the test.

The second step is selecting the content areas or topics from which test items will be drawn.

The third step is deciding on the cognitive levels to be assessed, such as knowledge, comprehension, or application.

The fourth step is allocating the number of items to each topic and cognitive level to ensure balanced coverage.

8. Explain four uses of statistics in educational research.

Statistics are used to summarize data into meaningful forms such as averages, percentages, and graphs.

They help in making comparisons between different groups or variables in a study.

They are essential in testing hypotheses to determine whether observed differences are significant.

They help interpret research results and draw conclusions that can guide educational practices.

9. Give two reasons why standard deviation is considered a better measure of variability than range.

Standard deviation uses all data points in its calculation, providing a more accurate picture of score distribution.

It is less affected by extreme scores than the range, making it a more reliable measure in most datasets.

10. (a) Define construct validity.

Construct validity refers to how well a test measures the theoretical concept or trait it is intended to assess.

It is concerned with whether the test truly reflects the underlying skills or abilities that it claims to measure.

(b) Differentiate between predictive validity and concurrent validity with examples.

Predictive validity measures how well a test can forecast future performance. For example, an entrance exam predicting a student's success in university studies.

Concurrent validity measures how well a test correlates with another established measure taken at the same time. For example, a new reading test being compared to an existing standardized reading assessment given in the same week.

11. (a) The following are the Physics test scores of 12 students: 45, 55, 60, 70, 65, 50, 75, 80, 55, 60, 85, 90.

(i) Calculate the mean score (nearest whole number).

Step 1: Sum of all scores = 45 + 55 + 60 + 70 + 65 + 50 + 75 + 80 + 55 + 60 + 85 + 90 = 790

Step 2: Mean = Total sum ÷ Number of students = 790 ÷ 12 = 65.83 ≈ 66

(ii) Calculate the variance and standard deviation (nearest whole number).

Step 1: Find the deviation of each score from the mean and square it:
$(45 - 66)^2 = 441$
$(55 - 66)^2 = 121$
$(60 - 66)^2 = 36$
$(70 - 66)^2 = 16$

$(65 - 66)^2 = 1$

$(50 - 66)^2 = 256$

$(75 - 66)^2 = 81$

$(80 - 66)^2 = 196$

$(55 - 66)^2 = 121$

$(60 - 66)^2 = 36$

$(85 - 66)^2 = 361$

$(90 - 66)^2 = 576$

Step 2: Sum of squared deviations = 2242

Step 3: Variance = $2242 \div 12 = 186.83 \approx 187$

Step 4: Standard deviation = $\sqrt{187} \approx 13.67 \approx 14$

(iii) Using a mean of 60 and a standard deviation of 10, standardize the scores of the student with the highest and lowest marks.

Highest score: X = 90

$Z = (90 - 66) \div 14 = 24 \div 14 \approx 1.71$

Standard score = $(1.71 \times 10) + 60 = 17.1 + 60 = 77.1 \approx 77$

Lowest score: X = 45

$Z = (45 - 66) \div 14 = -21 \div 14 \approx -1.5$

Standard score = $(-1.5 \times 10) + 60 = -15 + 60 = 45$

(b) Explain three reasons why a researcher might prefer using standard scores in interpreting test results.

Standard scores allow easy comparison of results across different tests that have varying scales.

They remove the effects of differences in test difficulty, making interpretation fairer.

They help identify how far a score is above or below the mean, making performance interpretation clearer.

12. A researcher plans to study the relationship between teachers' workload and students' academic performance in secondary schools.

(a) Formulate one broad research objective and three specific objectives for this study.

Broad objective: To examine the effect of teachers' workload on students' academic performance in secondary schools.

Specific objective 1: To determine how the number of teaching hours per week influences student achievement.

Specific objective 2: To analyze the relationship between the number of subjects taught by a teacher and student performance.

Specific objective 3: To assess whether class size impacts the quality of instruction and student achievement.

(b) Discuss four possible sources of bias in collecting data for this research and explain how each can be minimized.

Sampling bias can occur if the schools chosen are not representative. This can be minimized by using random selection.

Response bias may happen if respondents give socially desirable answers. This can be reduced by ensuring confidentiality.

Measurement bias can occur if the instruments used are unclear or unreliable. This can be minimized by piloting and refining them.

Researcher bias can happen if the researcher's expectations influence data interpretation. This can be minimized by using standardized procedures and double-checking data.

13. Critically analyze four limitations of using correlation coefficients in educational research, and explain how these limitations can be addressed.

Correlation does not indicate causation, so even a strong correlation cannot prove one variable causes the other. This can be addressed by supplementing with experimental research.

Correlation values can be affected by outliers, which may distort the strength of the relationship. This can be managed by screening and adjusting data.

Correlation assumes a linear relationship, which means it may miss strong but non-linear associations. This can be addressed by using scatter plots to check relationship shapes.

Small sample sizes can lead to unreliable correlation results. This can be addressed by using a sufficiently large and representative sample.

14. (a) Propose four different methods that could be used to estimate the reliability of a mathematics achievement test, explaining the conditions under which each method is most suitable.

Test-retest method involves giving the same test twice to the same group and is suitable when the construct being measured is stable over time.

Split-half method divides the test into two equal parts and checks consistency, suitable for tests measuring a single skill.

Parallel forms method uses two equivalent tests to the same group, suitable when alternate versions of a test are available.

Internal consistency method, such as Cronbach's alpha, measures the relationship between items, suitable when the test contains multiple items assessing the same construct.

(b) Suggest four ways the test developers could improve its reliability based on the results of the reliability estimates.

Increase the number of high-quality items to reduce random errors.

Ensure clear and unambiguous wording of all test items.

Standardize administration procedures so that conditions are uniform for all examinees.

Remove or revise items that show low correlation with the overall test score.

15. A government education board wants to introduce national performance assessments in all primary schools. Discuss the potential benefits and challenges of implementing such assessments, providing at least four points for each.

One benefit is that they provide consistent data for comparing performance across schools and regions.

They help identify areas where the curriculum or teaching methods need improvement.

They can guide policy decisions by showing trends in student performance over time.

They can motivate schools to improve teaching and learning quality.

One challenge is the cost of developing, administering, and marking large-scale assessments.

They may lead to teaching to the test, reducing focus on broader learning.

They may disadvantage schools with fewer resources.

They can create stress for young learners, affecting performance and well-being.

16. Examine four key ethical considerations in conducting educational research involving children, and explain how each can be practically implemented in a Tanzanian school context.

Informed consent must be obtained from both parents and children, which can be implemented through written forms explained in Swahili and English.

Confidentiality must be maintained by using codes instead of names in data records.

Protection from harm requires ensuring that participation does not cause physical, emotional, or academic stress.

Right to withdraw must be respected, with clear communication that participation is voluntary and without consequences.