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: 2002 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 validity in educational measurement.

Validity in educational measurement refers to the degree to which a test or assessment actually measures what it is intended to measure, and not something else. This ensures the results are meaningful and serve the purpose for which the test was created.

It is an essential quality because without validity, even if a test is reliable, the decisions made based on the results may be wrong.

For example, if a mathematics test contains too many reading comprehension questions, it may measure reading ability more than mathematical skill, reducing its validity.

A valid test should align directly with the learning objectives, the syllabus content, and the skills or knowledge that educators aim to assess.

2. Mention four characteristics of a good research problem.

A good research problem must be clear and specific so that anyone reading it understands exactly what will be studied. Ambiguous or vague wording can confuse the direction of the study.

It must be researchable, meaning it can be investigated using scientific methods, data collection, and analysis within the resources available to the researcher.

It should be significant, addressing an issue that is important to education, fills a gap in knowledge, or solves a real-world problem.

It must be feasible, considering the time available, budget, researcher skills, and access to participants or data sources.

3. State four advantages of using stratified random sampling in educational research.

It ensures representation from all key subgroups in the population, such as boys and girls, rural and urban schools, or different grade levels, reducing bias in results.

It increases the accuracy and reliability of results because each subgroup is proportionally represented, making the sample more like the whole population.

It allows researchers to analyze differences between subgroups in more detail, providing richer and more precise findings.

It reduces sampling error compared to simple random sampling when the population has distinct groups with different characteristics.

4. Give four limitations of using observation as a research method.

Observation can be time-consuming because the researcher must watch participants for extended periods to gather enough information.

It is prone to observer bias, where the researcher's personal expectations or assumptions influence what is

noticed and recorded.
Some important behaviors may happen rarely or unexpectedly, meaning the researcher might miss them during the observation period.
Participants may change their behavior when they know they are being watched, which can lead to unnatural or inaccurate data (Hawthorne effect).

5. List three functions of educational assessment in the learning process.

It measures learners' achievement, showing how well they have mastered the content and skills outlined in the curriculum.
It provides feedback to both teachers and learners, helping to improve teaching strategies and learner study habits.
It helps identify students' strengths and weaknesses, making it possible to plan remedial lessons for weaker areas and enrichment activities for stronger students.

6. State four qualities of a standardized test.

A standardized test is administered under the same conditions for all test-takers, ensuring fairness.
It uses consistent scoring procedures so that results are comparable across individuals and groups.
It is developed using professional test construction processes, such as item analysis and piloting, to ensure validity and reliability.
It has established norms, meaning scores can be compared to a reference group to interpret performance levels.

7. Mention four ways of improving the reliability of an achievement test.

Increase the number of high-quality items that measure the same construct to reduce the effect of random errors.
Use clear, simple, and precise language so that all students interpret questions in the same way.
Standardize the administration process, ensuring all students take the test under similar conditions with equal time and instructions.
Analyze the test after administration to identify and revise items that do not discriminate well between stronger and weaker students.

8. Explain four purposes of administering a pilot study before the main research.

A pilot study helps identify weaknesses or unclear items in research instruments so they can be improved before the main study.
It tests the feasibility of the entire research process, including timing, administration, and logistics, to

avoid problems later.
It gives an idea of how long it will take to collect and analyze data, which helps in scheduling and resource allocation.
It provides an opportunity to train research assistants, ensuring they understand procedures and can apply them consistently in the main study.

9. Give two reasons why range is considered a weaker measure of variability compared to standard deviation.

Range only considers the highest and lowest values in a dataset, ignoring all other scores, so it does not reflect the full spread of data.
It is highly affected by extreme values (outliers), which can make the range misleading as a representation of the overall variability.

10. (a) Define face validity and content validity.

Face validity is the extent to which a test appears to measure what it claims to measure, based on a quick judgment by test-takers or observers. It is more about appearance than scientific proof.
Content validity is the extent to which a test covers all the important topics, skills, and objectives it is supposed to measure, based on systematic expert review.

(b) Explain four differences between the two, with examples.

Face validity is subjective and based on first impressions, while content validity is objective and based on detailed expert analysis of test content.
Face validity focuses on how relevant or appropriate the test looks, whereas content validity examines whether the test actually represents the full scope of the subject matter.
Face validity can be judged quickly without deep analysis, but content validity requires comparing each test item to the syllabus and learning objectives.
For example, a science test that looks like it contains experiments may have face validity, but without matching the curriculum topics, it would lack content validity.

11. (a) The following are Mathematics scores for 16 students:

40, 45, 48, 50, 52, 55, 58, 60, 62, 65, 68, 70, 72, 75, 78, 80

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

The first step is to add all the scores together to get the total. The sum of the scores is 978.

The next step is to divide the total sum by the number of students to get the mean. 978 divided by 16 equals 61.125.

When rounded to the nearest whole number, the mean score is 61.

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

To find variance, begin by finding the deviation of each score from the mean, then square each deviation to remove negative signs.

The squared deviations for each score are then summed together, giving a total of 2328.

The variance is found by dividing this total by the number of students, giving 145.5, which rounds to 146.

The standard deviation is the square root of the variance, which is $\sqrt{146} \approx 12.08$, rounded to 12.

(iii) Using a mean of 65 and a standard deviation of 12, compute the z-scores for the highest and lowest scores.

For the highest score of 80, subtract the mean (65) to get 15, then divide by the standard deviation (12) to get a z-score of approximately 1.25.

For the lowest score of 40, subtract the mean (65) to get −25, then divide by the standard deviation (12) to get a z-score of approximately −2.08.

(iv) Interpret the two z-scores in the context of performance.

A z-score of 1.25 means the highest scorer performed 1.25 standard deviations above the mean, showing a level of performance clearly above the group average.

A z-score of −2.08 means the lowest scorer performed 2.08 standard deviations below the mean, indicating performance far lower than most of the group.

(b) State three advantages of using z-scores when reporting examination results.

Z-scores allow results from different tests to be compared fairly, even when the tests have different score ranges.

They make it easy to identify whether a score is above or below the average and by how much.

They highlight extreme performances, whether exceptionally high or low, which can help guide intervention or recognition.

12. A study is being conducted to examine the relationship between students' study habits and academic performance in Tanzanian secondary schools.

(a) State one general objective and three specific objectives for this study.

The general objective is to investigate the relationship between study habits and academic performance among secondary school students.

One specific objective is to determine the extent to which time management in studying affects student performance.

Another specific objective is to analyze how the use of different study techniques influences academic achievement.

A further specific objective is to examine the role of the study environment on students' academic results.

(b) Identify four possible sources of bias in this study and explain how each can be minimized.

Self-report bias could occur if students give exaggerated accounts of their study habits. This can be minimized by comparing their responses with teacher observations.

Sampling bias might happen if only high-performing schools are included in the study. This can be avoided by selecting schools randomly from different performance levels.

Measurement bias may arise if the questionnaire has unclear questions. This can be reduced by piloting and refining the instrument before the main study.

Researcher bias could occur if the researcher's personal expectations affect data interpretation. This can be minimized by using clear coding rules and involving multiple coders.

13. Explain four factors that a researcher should consider when deciding on the appropriate sample size for an educational research study.

The size of the population is a major factor. A larger population will often require a bigger sample to represent it accurately.

The desired level of precision affects sample size because higher accuracy in estimates requires more participants.

The variability of the population must be considered, as a population with more differences between individuals needs a larger sample to capture the range of characteristics.

The available resources, such as time, finances, and manpower, also determine how large the sample can be without exceeding practical limits.

14. A newly developed national examination has produced significantly different results when administered in urban and rural schools.

(a) Suggest four possible reasons for the discrepancy.

There may be differences in the quality of teaching resources available between urban and rural schools.

Access to learning materials such as textbooks, technology, and libraries may be higher in urban schools.

Socio-economic differences could affect students' preparation and learning opportunities in rural areas.

Environmental factors like larger class sizes or inadequate school infrastructure could hinder rural students' learning.

(b) Propose four strategies to ensure fairness in future administrations.

Provide equal access to high-quality teaching and learning resources in all schools, regardless of location.

Standardize teacher training across all regions to ensure similar quality of instruction.

Develop test items that are free from cultural or contextual bias that might favor urban students.

Conduct pre-testing of exam papers to identify and address any elements that might disadvantage specific groups.

15. Discuss four advantages and four disadvantages of using high-stakes testing in the Tanzanian education system.

One advantage is that it motivates students to work hard since important opportunities depend on the results.

Another advantage is that it provides clear benchmarks for comparing student performance across schools and regions.

It can be used to hold schools accountable for maintaining good teaching standards.

It helps in identifying top-performing students for scholarships, awards, and further educational opportunities.

One disadvantage is that it can cause stress and anxiety among students, which may reduce performance.

It may lead teachers to teach only the content that will be tested, neglecting broader educational skills.

It can disadvantage students from schools with fewer resources, leading to unfair comparisons.

It may encourage cheating or unethical behavior in an attempt to secure higher scores.

16. Examine four ethical issues that may arise when collecting data from minors in educational research, and suggest measures to address each.

Informed consent is essential, meaning researchers must get permission from both parents and students after explaining the study. This ensures participation is voluntary and informed.

Confidentiality must be maintained so that students' identities are protected. This can be done by using codes instead of names in all records.

Protection from harm must be guaranteed, ensuring that participation in the study does not cause any physical, emotional, or academic disadvantage.

The right to withdraw must be respected, meaning participants should be told they can leave the study at any time without facing any consequences.