THE UNITED REPUBLIC OF TANZANIA

NATIONAL EXAMINATIONS COUNCIL OF TANZANIA

DIPLOMA IN SECONDARY EDUCATION EXAMINATION

762 EDUCATIONAL RESEARCH, MEASUREMENT AND EVALUATION

Time: 3 Hours **ANSWERS** **Year: 2011**

Instructions

1. This paper consists of section A and B.

2. Answer all questions in section A, and four questions from section B.

SECTION A (40 Marks)

Answer all questions in this section.

1. (a) Mention two major types of test items. (b) State four reasons of administering a test to students.

(a) Major Types - Multiple Choice: One type is multiple-choice items, offering options. Science variety enhances teaching effectiveness and stability through diverse learning and educational outcomes in classrooms.

(a) Major Types - Essay: Essay items require detailed responses. Science depth improves teaching precision and stability through expressive education and learning strategies in instruction.

(b) Reason - Assessment: One reason is to assess student achievement. Science evaluation boosts teaching quality and stability through measured learning and educational progress in classrooms.

(b) Reason - Feedback: Feedback on performance is a reason. Science input enhances teaching impact and stability through responsive education and educational outcomes in learning environments.

(b) Reason - Diagnosis: Diagnosis of learning gaps is a reason. Science identification improves teaching reliability and stability through targeted learning and educational strategies in classrooms.

(b) Reason - Motivation: Motivation through goal setting is a reason. Science encouragement enhances teaching precision and stability through engaged education and learning tools in instruction.

2. List down four instruments used in assessing students' achievement with two columns and insert four differences between quantitative research and qualitative research.

Instruments - Tests: One instrument is tests, measuring knowledge. Science assessment enhances teaching effectiveness and stability through evaluated learning and educational outcomes in classrooms.

Instruments - Questionnaires: Questionnaires gather data. Science feedback improves teaching precision and stability through collected education and learning strategies in instruction.

Instruments - Observations: Observations assess behavior. Science monitoring boosts teaching quality and stability through observed learning and educational progress in classrooms.

Instruments - Interviews: Interviews provide insights. Science dialogue enhances teaching impact and stability through in-depth education and educational outcomes in learning environments.

Differences - Approach: Quantitative uses numerical data, qualitative uses descriptive data. Science method enhances teaching reliability and stability through varied learning and educational strategies in classrooms.

Differences - Objective: Quantitative is objective, qualitative is subjective. Science focus improves teaching precision and stability through balanced education and learning tools in instruction.

Differences - Sample Size: Quantitative uses large samples, qualitative uses small samples. Science scale boosts teaching quality and stability through representative learning and educational progress in classrooms.

Differences - Analysis: Quantitative uses statistics, qualitative uses themes. Science technique enhances teaching effectiveness and stability through analytical education and educational outcomes in learning environments.

3. Draw a table with research and qualitative research in each column.

Aspect

Quantitative Research

Qualitative Research

Data Type

Numerical data

Descriptive data

Objective

Objective measurement

Subjective interpretation

Sample Size

Large samples

Small samples

Analysis Method

Statistical analysis

Thematic analysis

Science comparison enhances teaching effectiveness and stability through structured learning and educational outcomes in classrooms.

4. Outline three strategies you may use for best construction of essay questions.

Clear Instructions: One strategy is providing clear instructions. Science clarity enhances teaching effectiveness and stability through guided learning and educational outcomes in classrooms.

Balanced Scope: Balanced scope avoids overly broad topics. Science focus improves teaching precision and stability through manageable education and learning strategies in instruction.

Rubric Development: Rubric development ensures fair marking. Science criteria boosts teaching quality and stability through consistent learning and educational progress in classrooms.

5. Distinguish between measurement and test by basing your responses on the meanings the terms concerned.

Definition - Measurement: Measurement quantifies attributes or performance. Science quantification enhances teaching effectiveness and stability through data-based learning and educational outcomes in classrooms. Test is a specific tool to assess knowledge, improving teaching precision and stability through evaluated education and learning strategies in instruction.

Purpose - Measurement: Measurement provides broad data, while test targets specific skills. Science intent boosts teaching quality and stability through purposeful learning and educational progress in classrooms.

Application - Measurement: Measurement is ongoing, test is periodic. Science timing enhances teaching impact and stability through continuous education and educational outcomes in learning environments.

6. Briefly explain two basic ways of interpreting student performance. Your explanation should focus on their basic functions.

Norm-Referenced Interpretation: One way is norm-referenced, comparing to peers. Science ranking enhances teaching effectiveness and stability through comparative learning and educational outcomes in classrooms. Its function is to identify relative standing.

Criterion-Referenced Interpretation: Criterion-referenced compares to standards. Science benchmark improves teaching precision and stability through standard-based education and learning strategies in instruction. Its function is to measure mastery.

7. What is the importance of research proposal. State four points.

Planning: One point is planning the research process. Science organization enhances teaching effectiveness and stability through structured learning and educational outcomes in classrooms.

Funding: Securing funding is a point. Science support improves teaching precision and stability through resourced education and learning strategies in instruction.

Guidance: Guidance for methodology is a point. Science direction boosts teaching quality and stability through guided learning and educational progress in classrooms.

Evaluation: Evaluation of feasibility is a point. Science assessment enhances teaching impact and stability through practical education and educational outcomes in learning environments.

8. Give four differences between basic research and applied research.

Purpose: One difference is purpose, with basic research seeking knowledge. Science exploration enhances teaching effectiveness and stability through theoretical learning and educational outcomes in classrooms. Applied research solves practical problems, improving teaching precision and stability through applied education and learning strategies in instruction.

Application: Basic is theoretical, applied is practical. Science focus boosts teaching quality and stability through relevant learning and educational progress in classrooms.

Outcome: Basic contributes to theory, applied to practice. Science result enhances teaching impact and stability through impactful education and educational outcomes in learning environments.

Funding: Basic often academic, applied industry-driven. Science source improves teaching reliability and stability through supported learning and educational strategies in classrooms.

9. (a) "A test may be reliable but need not be valid". Provide one basic reason to justify the statement. (b) Mention any two basic qualities of a good test.

(a) Justification - Consistency vs. Accuracy: One reason is that a test can be reliable (consistent) but not valid (accurate) if it measures the wrong thing consistently. Science distinction enhances teaching effectiveness and stability through precise learning and educational outcomes in classrooms. Example: A broken ruler measuring length the same way each time.

(b) Quality - Reliability: One quality is reliability, ensuring consistent results. Science repeatability improves teaching precision and stability through dependable education and learning strategies in instruction.

(b) Quality - Validity: Validity ensures the test measures intended outcomes. Science accuracy boosts teaching quality and stability through relevant learning and educational progress in classrooms.

10. Write four reasons as to why teachers should use formative evaluation during the teaching and learning process.

Feedback: One reason is providing feedback to students. Science input enhances teaching effectiveness and stability through responsive learning and educational outcomes in classrooms.

Progress Monitoring: Monitoring progress is a reason. Science tracking improves teaching precision and stability through assessed education and learning strategies in instruction.

Adjustment: Adjustment of teaching methods is a reason. Science adaptation boosts teaching quality and stability through flexible learning and educational progress in classrooms.

Motivation: Motivation through recognition is a reason. Science encouragement enhances teaching impact and stability through engaged education and educational outcomes in learning environments.

SECTION B (60 Marks)

Answer four (4) questions from this section. Question 11 is compulsory.

11. Study carefully the scores of 16 students obtained from Curriculum and Teaching test and answer the following questions:

(a) Find: (i) Mode (ii) Median (iii) Range

(i) Mode: The mode is 58, appearing twice. Science frequency enhances teaching effectiveness and stability through common learning and educational outcomes in classrooms.

(ii) Median: Ordered scores: 29, 32, 37, 40, 46, 50, 55, 56, 58, 58, 60, 75, 80, 88, 90. Median is the 8th value: 58. Science midpoint improves teaching precision and stability through central education and learning strategies in instruction.

(iii) Range: Range is the difference between the highest (90) and lowest (29): 90 - 29 = 61. Science spread boosts teaching quality and stability through variability learning and educational progress in classrooms.

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

Standard Deviation: Steps: Mean = (29 + 32 + 37 + 40 + 46 + 50 + 55 + 56 + 58 + 58 + 60 + 75 + 80 + 88 + 90) / 15 ≈ 56.67. Variance ≈ 392.44, Standard Deviation = √392.44 ≈ 19.81 ≈ 20. Science dispersion improves teaching reliability and stability through statistical learning and educational strategies in classrooms.

(c) Using Z-score and T-score formulae, standardize the scores for student J and P by using z-score and t-score formulae.

Mean and SD: Mean ≈ 56.67, SD ≈ 19.81.

(c) Z-Score for J (80): Z = (80 - 56.67) / 19.81 ≈ 1.18. Science standardization enhances teaching precision and stability through comparative education and learning tools in instruction.

(c) T-Score for J: T = 50 + 10 x 1.18 ≈ 61.8 ≈ 62. Science scaling boosts teaching quality and stability through adjusted learning and educational progress in classrooms.

(c) Z-Score for P (Note: No score for P, assuming O with 88): Z = (88 - 56.67) / 19.81 ≈ 1.58. Science adjustment enhances teaching impact and stability through relative education and educational outcomes in learning environments.

(c) T-Score for P/O: T = 50 + 10 x 1.58 ≈ 65.8 ≈ 66. Science transformation improves teaching reliability and stability through scaled learning and educational strategies in classrooms.

12. By providing six points, suggest how multiple-choice items can best be constructed.

Multiple-choice items are commonly used in assessments because they allow for objective grading and broad content coverage. To ensure their effectiveness, they must be carefully designed.

One important aspect of constructing multiple-choice items is writing clear and concise questions. The question (stem) should be free of unnecessary information and should clearly state the problem. Avoiding ambiguous wording helps ensure that students focus on the actual concept being tested rather than struggling to interpret the question.

Another key point is ensuring that there is only one correct answer. Distractors (incorrect options) should be plausible but clearly incorrect for students who do not know the correct response. This prevents students from guessing correctly based on poorly designed choices.

The options should be of similar length and structure. If the correct answer is significantly longer or more detailed than the distractors, students may guess the answer simply based on length rather than knowledge. Keeping all answer choices similar in style prevents test-taking strategies from influencing the results.

Using negative wording, such as "Which of the following is NOT correct?" should be avoided unless necessary. Negative questions can be confusing and may mislead students into selecting the wrong answer simply due to misinterpretation. If negatives must be used, they should be clearly emphasized, such as by capitalizing "NOT."

The placement of the correct answer should be randomized. If the correct answers tend to follow a pattern (e.g., mostly "C"), students may identify the pattern rather than applying their knowledge. Randomizing the order of the correct answers ensures a fair test.

Finally, multiple-choice items should assess understanding rather than simple recall. Instead of testing basic memorization, well-constructed questions should require students to apply concepts, analyze information, or solve problems. This improves the quality of assessment and ensures that students' critical thinking skills are being evaluated.

13. Explain how the test-retest method is conducted.

The test-retest method is used to measure the reliability of a test by administering the same test to the same group of individuals at two different points in time. This method assesses the consistency of the test results over time.

First, the test is given to a group of participants under standard conditions. Their responses are recorded, and their scores are calculated. This initial administration serves as the baseline measurement.

After a period of time, usually days or weeks, the same test is given to the same group again under similar conditions. The interval between the two test administrations should be long enough to prevent recall of previous answers but short enough to ensure that no significant changes in the test-takers' knowledge or abilities have occurred.

The scores from both test administrations are then compared. If the results are highly similar or correlated, the test is considered reliable, meaning it produces consistent results over time. A high correlation coefficient (close to 1) indicates strong reliability, while a low correlation suggests that the test may not be consistent.

This method is commonly used in psychological assessments, educational testing, and research studies to ensure that test scores remain stable over time and are not influenced by external factors.

14. Explain four disadvantages of the test-retest method.

The test-retest method, while useful for measuring reliability, has several limitations that may affect its accuracy and practicality.

One disadvantage is the possibility of memory effects. If the time gap between test administrations is too short, participants may remember their previous answers, leading to artificially high reliability scores. This does not truly reflect the consistency of the test but rather the ability of participants to recall previous responses.

Another issue is the potential for changes in the test-taker's ability or condition. If too much time passes between the two test administrations, external factors such as learning, fatigue, or environmental influences may alter the individual's performance. This makes it difficult to determine whether differences in scores are due to test inconsistency or genuine changes in the participant's abilities.

The test-retest method can also be affected by test-taker motivation. A person may be highly motivated during the first test but less interested during the second attempt, leading to variations in performance that do not reflect the test's reliability. Factors like mood, health, or external distractions can impact how consistently a participant performs.

Finally, this method is time-consuming and requires additional resources. Administering the same test twice, ensuring consistent conditions, and analyzing the results require extra effort, making it impractical for large-scale assessments or time-sensitive research.

15. Compare and contrast questionnaire and interview by providing three explanations in each case.

Questionnaires and interviews are both methods of collecting data, but they differ in how information is gathered and the level of interaction with respondents.

One similarity is that both methods aim to collect information from individuals or groups to analyze opinions, behaviors, or experiences. They are widely used in research, education, and social studies to gather data efficiently.

Another similarity is that both can be structured, semi-structured, or unstructured. A structured format ensures that all respondents answer the same questions, while an unstructured format allows for flexibility in responses. Semi-structured approaches combine both methods to balance consistency and depth.

Both methods also require careful design to avoid bias. Whether through question phrasing in a questionnaire or interviewer influence in an interview, researchers must ensure that questions do not lead respondents toward a particular answer.

Despite these similarities, there are key differences between questionnaires and interviews. One major difference is the level of interaction. A questionnaire is typically a self-administered written document that respondents complete independently, while an interview involves direct communication between the

researcher and the participant. This allows interviews to provide deeper insights through follow-up questions.

Another difference is the level of detail in responses. Interviews allow for more detailed and open-ended answers since interviewers can probe further into responses. Questionnaires, on the other hand, are usually limited to predefined options or short written responses, which may not capture the full depth of a participant's perspective.

Finally, questionnaires are generally more efficient and cost-effective. Since they can be distributed to a large number of people simultaneously (especially online), they require less time and effort compared to interviews, which involve scheduling, conducting, and transcribing responses. However, interviews provide richer and more personalized data, making them useful for in-depth studies.

16. Describe each of the four types of scales of measurement in education and provide one example of each type.

In educational research and assessment, data is measured using four main types of scales: nominal, ordinal, interval, and ratio. Each type has distinct characteristics and is used for different kinds of analysis.

The nominal scale is the simplest form of measurement, used for categorizing data without any numerical or ranking order. It classifies data into distinct groups based on attributes. An example in education is classifying students based on their learning styles (e.g., visual, auditory, or kinesthetic). Since there is no inherent ranking, calculations like averages do not apply to nominal data.

The ordinal scale involves ranking data in a specific order but does not indicate the precise difference between ranks. This scale is commonly used when measuring subjective attributes like opinions or performance levels. For example, grading students as "Excellent," "Good," "Average," or "Poor" ranks their performance but does not specify the exact score differences between the categories.

The interval scale provides not only order but also equal differences between values. However, it lacks a true zero point, meaning ratios cannot be meaningfully calculated. A common example in education is standardized test scores, such as IQ scores. The difference between 90 and 100 IQ points is the same as the difference between 100 and 110, but an IQ of 0 does not indicate an absence of intelligence.

The ratio scale is the most advanced, as it includes all properties of the interval scale but also has a true zero point, allowing for meaningful ratio comparisons. In education, an example is students' test scores measured in percentages. A score of 80% is twice as high as a score of 40%, and a score of 0% represents a complete lack of achievement.

These four measurement scales are essential in educational research as they determine how data is collected, analyzed, and interpreted.

17. Analyze six characteristics of subjective type of test items.

Subjective test items are those that require students to construct their own responses rather than selecting from given options. These types of test items have several characteristics that define their nature and application in assessments.

One key characteristic is the necessity for high-level cognitive skills. Subjective test items demand more than just memorization; they require critical thinking, synthesis, and the ability to articulate thoughts effectively. For example, essay questions require students to analyze and evaluate information before forming a coherent response.

Another characteristic is their openness to multiple interpretations. Unlike objective tests, which have a single correct answer, subjective test items can have varied but equally valid responses. This allows for flexibility in assessing students' understanding but can also introduce inconsistency in grading.

Subjective test items also require a higher level of language proficiency. Since responses must be articulated in a structured manner, students who struggle with language skills might face challenges even if they understand the content well. This can be a limitation for learners whose first language is different from the language of instruction.

A major characteristic is the time-consuming nature of grading. Unlike multiple-choice questions, which can be marked quickly, subjective responses require careful reading, interpretation, and evaluation. This makes grading more labor-intensive and subject to human bias.

Subjective test items provide deeper insight into students' thought processes. They allow educators to assess how well students can apply knowledge, make connections between concepts, and present reasoned arguments, making them particularly useful in fields that require critical thinking and creativity.

Finally, subjective test items often include a degree of subjectivity in scoring. Even with a clear rubric, different evaluators may interpret and grade responses differently. This variability can affect consistency in assessment outcomes, making it necessary for teachers to use detailed rubrics to ensure fairness.

18. A teacher wants to stream (band) his 10 pupils after obtaining test scores from the English, Mathematics, and Science tests. Examine the given scores in Table 1 below and answer the following questions:

(a) By using a ruler and a pencil, or a free-hand drawing, draw Table 2 in your answer booklet.

This question requires the student to manually recreate Table 2 as provided in the second image. The table should include the columns for pupil names, standardized scores for English, Mathematics, and Science, total scores, and new class positions after standardization. This step is important for organizing the standardized data before further analysis.

(b) By using z-score given in Table 1, use T-score to standardize English, Mathematics and Science scores and insert them in Table 2 which you have just drawn.

To standardize the scores using the T-score formula, the following formula is applied:

$T = 50 + (10 \times \text{z-score})$

For each subject, the z-scores given in Table 1 must be substituted into this formula to compute the T-scores. These standardized scores should then be recorded in the corresponding columns of Table 2. The T-score transformation allows for better comparison of performance across different subjects.

(c) Find the total score of every pupil and insert them in Table 2.

After computing the T-scores for English, Mathematics, and Science, the total score for each student is determined by summing their T-scores across the three subjects. This total score should be added to the "Total Score" column in Table 2.

(d) Arrange the pupils' names by following the given new position numbers (starting from 1 – 10).

Once the total standardized scores have been calculated, the pupils should be ranked in descending order. The student with the highest total score receives the first position, while the one with the lowest total score is placed last. This step reorganizes the students' rankings based on their standardized performance.

(e) What is the new position of pupil A in the group?

Pupil A's new position is determined by comparing their total standardized score with those of the other students. If pupil A initially had a high total score but their standardized total is lower, their ranking may drop. The final position can be identified after arranging all students' standardized scores in descending order.

(f) Give a major reason that made pupil A's first position in Table 1 to change to the present position in Table 2 after standardization had taken place.

The change in pupil A's position is mainly due to the impact of standardization. Standardization adjusts raw scores based on the distribution of the dataset, accounting for variations in difficulty across subjects. If pupil A had higher raw scores but performed in subjects with lower variation, their advantage may be diminished after applying T-scores. Conversely, students who performed well in subjects with high variability could benefit more from standardization, leading to shifts in ranking.