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: 2009**

Instructions

1. This paper consists of section A and B.

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

1. (a) Define reliability of a test.

Reliability of a test refers to the consistency of the test results when it is administered multiple times under the same conditions. A reliable test produces similar results for the same individuals over repeated attempts, assuming no significant changes in their abilities or knowledge. For example, if a student takes the same test twice within a short period without any additional learning, a reliable test should yield nearly identical scores.

(b) Which three (3) characteristics can reduce the reliability of a test considerably?

One characteristic that reduces reliability is inconsistent test administration. If the conditions under which the test is given vary—such as changes in instructions, timing, or environmental distractions—then students may perform differently based on these external factors rather than their true abilities.

Another characteristic that negatively impacts reliability is poor question design. Ambiguous, vague, or misleading questions can cause students to interpret them differently, leading to inconsistent answers. A test must be carefully designed so that every test-taker understands the questions in the same way.

The third factor that can lower reliability is subjective scoring. When grading criteria are unclear, or when different examiners interpret responses differently, test results can be inconsistent. Tests with essay-type questions, for example, require detailed rubrics to ensure all students are graded fairly and consistently.

2. (a) Define validity of a test.

Validity of a test refers to the extent to which the test measures what it is intended to measure. If a mathematics test includes questions on history, it would lack validity because it is not assessing mathematical ability. A valid test ensures that the questions align with the subject matter and accurately reflect the skills or knowledge being tested.

(b) Which three (3) characteristics can reduce the validity of a test considerably?

One major factor that reduces validity is poor alignment with learning objectives. If a test does not cover the intended subject areas, then it does not provide an accurate measure of student achievement. For example, if a science test includes unrelated math problems, it fails to measure students' science knowledge effectively.

Another issue that lowers validity is cultural or language bias. If test questions assume prior knowledge that only certain groups of students have, it can disadvantage others. For example, using complex English

vocabulary in a test meant for students who are still learning the language could lead to misinterpretation and lower scores, even if those students understand the subject.

A third factor that affects validity is the presence of external influences, such as test anxiety or unfamiliar test formats. If a student performs poorly due to stress or confusion about how to answer certain question types, their score may not reflect their actual knowledge or skills.

3. Tabulate the advantages and disadvantages of an interview as a research tool for data collection.

An interview as a research tool has both advantages and disadvantages. One advantage is that it allows for in-depth responses. Unlike surveys with fixed-choice answers, interviews enable respondents to explain their thoughts, which can provide richer and more detailed data.

Another advantage is that interviews allow for clarification. If a participant does not understand a question, the interviewer can provide explanations, ensuring that responses are accurate and meaningful.

Interviews are also flexible, meaning that the interviewer can adjust questions based on the responses given. This adaptability makes interviews useful for exploring topics that require deeper understanding.

However, interviews have disadvantages as well. One major disadvantage is that they are time-consuming. Conducting one-on-one interviews with multiple participants takes much longer than distributing and collecting survey forms.

Another disadvantage is the potential for interviewer bias. The way an interviewer asks questions or reacts to responses can unintentionally influence the answers given by participants, which can compromise the objectivity of the data collected.

A final disadvantage is the cost. Interviews require resources such as recording equipment, transcription services, and travel expenses, making them a more expensive data collection method compared to questionnaires.

4. Differentiate educational measurement from educational assessment.

Educational measurement is the process of assigning numerical values to a student's learning outcomes. It typically involves tests, quizzes, and standardized exams that produce scores or grades. For example, measuring a student's performance in mathematics by assigning a test score of 85% is an example of educational measurement.

Educational assessment, on the other hand, is a broader concept that includes both measurement and evaluation. It involves interpreting test results and using them to make decisions about teaching and learning. Assessment can include written feedback, self-reflection, and teacher observations in addition to test scores.

For example, a teacher assessing a student's writing skills may not only consider the test score but also provide feedback on grammar, coherence, and creativity.

5. Give a short description of each of the four (4) scales of measurement in education.

The nominal scale is the simplest level of measurement, where data is categorized without any specific order. For example, grouping students by their favorite subject (Math, Science, English) is a nominal classification because there is no ranking or numerical significance.

The ordinal scale arranges data in a meaningful sequence but does not specify the difference between ranks. For instance, if students are ranked as first, second, and third in a competition, we know their positions, but we do not know how much better one performed compared to the other.

The interval scale involves ordered categories with equal intervals between values but does not have a true zero point. A good example is temperature measured in Celsius or Fahrenheit. The difference between 20°C and 30°C is the same as the difference between 30°C and 40°C, but zero does not mean "no temperature."

The ratio scale is similar to the interval scale but includes a true zero, making it possible to compare values meaningfully. Examples include test scores and height measurements, where a score of 0 means the student has no correct answers.

6. Identify two (2) advantages and two (2) disadvantages of using an interview as a research tool for data collection.

One advantage of using interviews is that they provide deeper insights into participants' thoughts and opinions. Unlike multiple-choice surveys, interviews allow respondents to explain their reasoning, which can help researchers gather more meaningful data.

Another advantage is that interviews allow researchers to observe non-verbal cues, such as facial expressions and body language, which can provide additional context to the spoken responses.

However, one disadvantage is that interviews can be expensive and time-consuming, especially if many participants need to be interviewed. Conducting interviews, transcribing responses, and analyzing the data require significant effort and resources.

Another disadvantage is the risk of interviewer bias. The way an interviewer phrases questions or reacts to answers may unconsciously influence the respondent's replies, affecting the reliability of the data.

7. In what ways does a criterion-referenced measurement differ from a norm-referenced measurement?

A criterion-referenced measurement evaluates a student's performance based on a predetermined standard. For example, a driving test requires individuals to meet specific criteria, such as parking correctly or

following traffic rules. Their performance is judged based on whether they meet these standards rather than how they compare to others.

A norm-referenced measurement, on the other hand, compares a student's performance to that of a larger group. Standardized tests, such as IQ tests, use norm-referencing to rank students based on their relative performance rather than a fixed standard.

8. Give four (4) characteristics of a good hypothesis.

A good hypothesis should be clear and specific. It must state exactly what is being tested and avoid vague or broad statements. For example, "Increased study time improves test scores" is clearer than "Study habits affect grades."

It should be testable, meaning that it can be confirmed or disproven through research and data collection. If a hypothesis cannot be measured or observed, it is not useful in scientific studies.

A good hypothesis must be based on existing knowledge and literature. It should be developed from previous research and logical reasoning rather than random guesses.

It should also be predictive, meaning it establishes a relationship between variables and suggests a possible outcome. For example, "Students who receive daily reading practice will improve their vocabulary" predicts an effect of one factor on another.

9. What is the difference between educational measurement and educational evaluation?

Educational measurement focuses on assigning numerical values to student performance, such as test scores and grades. It is purely quantitative and does not involve interpretation.

Educational evaluation, however, is a broader process that includes analyzing and interpreting measurements to make informed decisions about teaching and learning strategies.

10. What is a standardized achievement test?

A standardized achievement test is a formal assessment designed to measure students' knowledge and skills under uniform conditions. It follows strict guidelines to ensure fairness, reliability, and comparability across different test-takers.

11. Convert the following raw scores obtained from a Science test into standard T-scores using a new standard deviation of 10 and a new mean of 50.

To convert raw scores into T-scores, we use the formula:

$T = 50 + 10 \times (X - \text{Mean}) / \text{Standard Deviation}$

The raw scores provided need their mean and standard deviation calculated first. Once we determine these values, we substitute them into the formula to obtain T-scores for each raw score. T-scores help standardize test results, making it easier to compare students' performance across different tests by placing all scores on the same scale.

(a) Prepare a Table of Specifications for a Civics test to cover specific topics.

A Table of Specifications (TOS) is a two-way chart that helps teachers align test questions with learning objectives. It ensures a balanced representation of topics and cognitive levels (such as knowledge, comprehension, application, and analysis). When preparing a TOS, we list topics in one column and cognitive domains in another. Then, we assign the number of questions to each area based on its importance in the curriculum. For instance, if democracy is a major topic, it may have more evaluation-based questions compared to a smaller topic like globalization.

(b) What is the total number of items in the test?

The total number of items in the test depends on the weight assigned to each topic and cognitive domain in the Table of Specifications. Teachers determine this based on curriculum coverage, time allocation, and assessment objectives. A well-balanced test should fairly distribute questions across all important topics.

(c) Which topic was given the greatest weighting?

The topic with the greatest weighting is usually the one considered most critical in the learning objectives. This is identified in the Table of Specifications by looking at the topic with the highest number of test items. Key topics often receive more weighting to reflect their importance in achieving the intended learning outcomes.

(d) Which levels of cognitive domain were given the greatest weighting?

Cognitive domains, such as knowledge, comprehension, application, and analysis, are essential in assessment design. The level given the greatest weighting depends on whether the test focuses more on recall-based knowledge or higher-order thinking skills. In many standardized tests, comprehension and application levels often receive the highest weighting to ensure students can use knowledge effectively rather than just memorize facts.

(e) Why should topics be given different weighting in the same test?

Topics should have different weightings to reflect their importance in the curriculum. Not all topics carry the same significance in achieving learning objectives, so more crucial concepts should have a higher proportion of questions. For example, in a science test, fundamental concepts like the laws of motion may require more questions than minor historical facts about scientists. This ensures a fair and comprehensive evaluation of students' understanding.

(f) Why is a Table of Specifications important in the construction of a standard test?

A Table of Specifications ensures that a test is valid and balanced by aligning assessment questions with learning objectives. Without a TOS, a test might overemphasize some topics while neglecting others, leading to an unfair evaluation. Additionally, it helps teachers create a variety of question types that assess different levels of understanding, from basic recall to critical thinking.

12. Justify by providing five points the importance of analyzing the students' responses to test items.

Analyzing students' responses to test items helps identify areas where students struggle, allowing teachers to adjust their teaching strategies. If a majority of students answer a question incorrectly, it may indicate a gap in understanding or an issue with the question itself.

Another reason is that response analysis helps improve test quality. By reviewing how students respond, educators can identify poorly worded or ambiguous questions that may confuse test-takers and adjust them for future assessments.

It also allows for the evaluation of curriculum effectiveness. If students consistently perform poorly in a particular subject area, it may indicate that the curriculum needs revision to address weaknesses.

Additionally, analyzing test responses helps in personalizing student learning. Teachers can use response patterns to provide targeted interventions, such as extra practice in weak areas or advanced challenges for students who excel.

Finally, response analysis supports fairness and equity in assessment. By studying patterns in responses, teachers can identify biases in test design and ensure that all students, regardless of background, have an equal opportunity to succeed.

13. Discuss four steps that a teacher should follow when planning a test.

The first step in planning a test is defining learning objectives. Teachers must determine what knowledge and skills the test should measure. This ensures that test questions align with curriculum goals and accurately assess student learning.

The second step is selecting the test format. Depending on the objectives, teachers choose between multiple-choice questions, short answers, essays, or practical assessments. The format should match the level of understanding being evaluated.

The third step is preparing a Table of Specifications. This step involves distributing test questions across different topics and cognitive levels to ensure a balanced assessment. This prevents overemphasis on certain areas while neglecting others.

The final step is reviewing and refining the test. Before administration, the teacher should check for errors, ambiguities, or biases in the questions. Piloting the test with a small group of students can help identify any issues that need correction.

14. Discuss characteristics of a good hypothesis.

A good hypothesis is clear and specific. It should precisely state the relationship between variables so that it is testable through research. Ambiguous hypotheses lead to confusion and make it difficult to draw conclusions.

It must be testable, meaning that it can be examined through experiments, surveys, or other research methods. If a hypothesis cannot be tested, it is not useful for scientific inquiry.

A strong hypothesis is based on existing knowledge. It should be formulated from previous research and logical reasoning, ensuring that it builds upon established findings rather than random assumptions.

It should also be predictive. A hypothesis should indicate what is expected to happen in a given situation. For example, "Students who study regularly will score higher on tests" makes a clear prediction that can be verified through research.

15. Discuss a strategy to employ elements of assessment, measurement, and evaluation in a classroom process.

One effective strategy is using formative assessment techniques, such as quizzes, observations, and student reflections, to gather data on learning progress. These assessments provide ongoing feedback that helps teachers adjust instruction.

Measurement comes into play by assigning numerical values to student performance through tests and graded assignments. This allows for quantifiable tracking of student progress over time.

Evaluation involves interpreting assessment data to make instructional decisions. Teachers analyze results to identify strengths and weaknesses, then modify teaching methods accordingly. For example, if many students struggle with a concept, the teacher may revisit the topic using a different approach.

16. Explain the advantages and limitations of an essay test.

One advantage of essay tests is that they assess critical thinking and the ability to express ideas clearly. Unlike multiple-choice questions, essays allow students to explain their reasoning and demonstrate deeper understanding.

Another advantage is flexibility. Essay questions can cover a wide range of topics, making them suitable for assessing complex subjects that require discussion and analysis.

However, a limitation of essay tests is subjectivity in grading. Different teachers may interpret responses differently, leading to inconsistent scores. Using clear rubrics can help reduce bias.

Another limitation is the time required for both students and teachers. Writing essays takes longer than answering objective questions, and grading them requires significant effort, especially for large classes.

17. Discuss the main characteristics of qualitative research and how it is applied in education.

Qualitative research focuses on exploring human experiences, behaviors, and social interactions in-depth rather than using numerical data. It often involves methods such as interviews, observations, and case studies.

One characteristic is its emphasis on context. Qualitative research considers the setting and background of participants to understand their perspectives fully. In education, this helps analyze classroom dynamics and teaching methods.

It is also flexible, allowing researchers to adjust their questions based on emerging insights. This makes it useful for studying complex educational issues, such as student motivation or teacher-student relationships.

18. Describe briefly ten of the main elements of an educational research report.

Title

The title of an educational research report clearly and concisely states the topic of the study. It should be specific enough to reflect the focus of the research while being broad enough to capture the reader's interest. A well-crafted title helps readers understand the scope of the research at a glance.

Abstract

The abstract provides a brief summary of the entire research report, usually within 150–250 words. It highlights the research problem, objectives, methodology, key findings, and conclusions. This section allows readers to quickly grasp the essence of the study and decide whether to read the full report.

Introduction

The introduction sets the stage for the research by explaining its background, significance, and purpose. It typically includes a statement of the research problem, research questions, and objectives. This section also justifies why the study is important and how it contributes to the field of education.

Literature Review

The literature review examines previous studies and theoretical frameworks relevant to the research topic. It helps establish the foundation for the study by identifying gaps in existing research and positioning the current study within the broader academic conversation. A strong literature review demonstrates the researcher's knowledge of the field and provides context for the study.

Methodology

The methodology section describes the research design, participants, data collection methods, and data analysis procedures. It provides enough detail so that other researchers can replicate the study if needed. This section ensures transparency and credibility by explaining how the research was conducted.

Results

The results section presents the findings of the study based on the data collected. It often includes tables, charts, or graphs to illustrate key trends and patterns. This section is purely focused on presenting the data without interpretation, allowing the findings to stand on their own.

Discussion

The discussion section interprets and analyzes the findings in relation to the research questions and existing literature. It explores possible explanations for the results, compares them with previous studies, and discusses their implications. This section often highlights the strengths and limitations of the study and suggests areas for future research.

Conclusion

The conclusion summarizes the key findings and their significance. It reinforces the main points of the study and connects them to the broader educational context. This section often provides recommendations for educators, policymakers, or researchers based on the study's findings.

References

The references section lists all the sources cited in the research report, following a specific citation style such as APA, MLA, or Chicago. Proper referencing gives credit to previous researchers and ensures academic integrity. A well-structured reference list allows readers to verify sources and further explore related studies.

Appendices

The appendices include additional materials that support the research but are not essential to the main text. These may include raw data, survey questionnaires, interview transcripts, or detailed statistical analyses. This section provides extra information for readers who want to examine the research in more detail.