Curriculum Development & Assessment Practice Test 8

60 original multiple-choice questions for teacher test preparation (FPSC / FGEI / PPSC style). Choose one answer for each question. The full answer key is on the last page — worked explanations for every question are free on the test page at QuaidianPoint.com.

Name: _____ Class: _____ Score: _____ / 60

Assessment Concepts and Purposes

Q1. A teacher keeps a portfolio of each student's drafts, projects and reflections across the term and reviews it with the student to plan the next steps. This practice best illustrates assessment as:

(A) A single final judgement
(B) A method of punishing incomplete work
(C) An ongoing, evidence-based process that supports learning decisions
(D) A replacement for the syllabus

Q2. Which statement best separates evaluation from measurement?

(A) Measurement records a score, while evaluation interprets that score against a purpose or standard to reach a judgement
(B) Measurement uses numbers, while evaluation uses only written comments
(C) Measurement is done at the end of term, while evaluation is done during teaching
(D) Measurement judges worth, while evaluation only counts correct answers

Q3. A school compares its students' science results with national benchmarks in order to decide whether to revise its science teaching. The assessment is being used for:

(A) Selecting students for a team
(B) Awarding individual certificates
(C) Checking daily homework completion
(D) Programme improvement based on evidence

Q4. Assessment for learning is most closely associated with:

(A) Final certification at the end of a course
(B) Feedback during learning that helps students close the gap between current and desired performance
(C) Ranking applicants for a scholarship
(D) Publishing school league tables

Q5. Which of the following is an example of formal assessment?

(A) A teacher's mental note during a class discussion
(B) A casual question asked while walking around the room
(C) A quick impression of a student's confidence
(D) A planned written test with a scoring key, given under the same conditions to the whole class

Q6. A teacher who relies only on one end-of-year examination for all decisions about students risks:

(A) Making the examination too easy for weaker students
(B) Making decisions from a narrow sample that may be affected by a single bad day
(C) Reducing the number of students who pass
(D) Increasing the time teachers spend marking

Q7. In assessment terminology, the "domain" of a test refers to:

(A) The full set of knowledge and skills the test is intended to sample, as defined by the syllabus and objectives
(B) The set of questions included in one particular test paper
(C) The level of difficulty of the items in the test
(D) The group of pupils whose scores are used to set the norms

Q8. Good classroom assessment practice requires that assessment tasks be:

(A) Secret from students until the examination day
(B) Unrelated to what was taught, to test general ability
(C) Aligned with the objectives taught and fair to all students
(D) Identical every year regardless of the syllabus

Formative, Summative, Diagnostic and Placement Assessment

Q9. During a lesson on fractions, the teacher asks students to show their working on mini whiteboards, spots a common error, and re-explains it immediately. This is:

(A) Formative assessment, because the evidence is used at once to adjust teaching
(B) Summative assessment, because marks are recorded
(C) Placement assessment, because groups are formed
(D) A final examination, because the whole class took part

Q10. Which feature distinguishes a summative test from a formative quiz on the same topic?

(A) The summative test covers less content than the formative quiz
(B) The formative quiz is given at the end of the topic to award a grade
(C) The summative test gives feedback that students can use to improve before grading
(D) The summative test is used to sum up achievement for a grade or decision, after learning is complete

Q11. A student who reads fluently aloud but cannot answer questions about a passage may need diagnostic assessment of:

(A) Spelling accuracy in written work
(B) Speed of reading aloud
(C) Reading comprehension, to locate where understanding breaks down
(D) Phonemic awareness for decoding new words

Q12. Diagnostic testing is most valuable when it leads to:

(A) A lower grade for the student
(B) A targeted remedial plan matched to the identified difficulty
(C) Moving the student to another school
(D) Repeating the same lesson unchanged

Q13. A new student joins a school mid-year. A short test of prior mathematics is given to decide which class and textbook level suit her. This decision is based on:

(A) A final examination
(B) Placement assessment
(C) A diagnostic study of a disability
(D) A summative report

Q14. Formative feedback is most effective when it:

(A) Only gives a grade with no comment
(B) Compares the student with the top student in class
(C) Describes what was done well, what needs improvement, and how to improve it
(D) Is given months after the work

Q15. A teacher gives a pre-test on a unit, finds most students already know half the content, and adjusts the starting point of the unit. This pre-test functioned mainly as:

(A) A summative examination
(B) A certification test
(C) A punishment for students who had not revised
(D) Placement and readiness assessment, showing where instruction should begin

Q16. Which of these is a summative use of assessment?

(A) Using annual board results to award a pass or fail grade for the year
(B) Checking understanding midway through a lesson
(C) Asking students to peer-review a draft before submission
(D) Giving a hint during a class activity

Q17. A formative assessment programme in a school is failing because teachers collect quiz data but never discuss it or change lessons. The missing element is:

(A) More frequent examinations
(B) Longer test papers
(C) Stricter marking
(D) The feedback loop: using evidence to adapt teaching and learning

Q18. Diagnostic assessment of a student's persistent spelling errors would most usefully examine:

(A) Patterns in the errors, such as particular letter combinations or rules, rather than only the total number wrong
(B) The total number of spelling errors in the term
(C) The student's reading speed during dictation
(D) How the student's spelling compares with the class average

Norm-Referenced and Criterion-Referenced Testing

Q19. A school reports that a student is in the top quarter of her year group in mathematics. This is a norm-referenced statement because it:

(A) Lists the topics she has mastered
(B) Locates her performance relative to her peers
(C) States the pass mark for the test
(D) Gives the number of items she answered

Q20. Criterion-referenced scores are especially useful for reporting to parents because they can show:

(A) The student's rank in the district
(B) How the student compares with the national average
(C) Which specific skills the child has and has not yet mastered
(D) The difficulty of the test paper

Q21. A teacher sets a mastery standard of 80 percent correct on a spelling unit test before the test is given. Students at or above that standard are judged to have mastered the unit. This standard-setting is characteristic of:

(A) Criterion-referenced testing
(B) Norm-referenced testing
(C) Diagnostic testing only
(D) Aptitude testing only

Q22. In a norm-referenced test, if the whole class performs very well, the relative ranking of students:

(A) Disappears completely
(B) Still spreads students out by comparison with one another, even though most have learned the material
(C) Becomes a measure of mastery
(D) Must be replaced by a pass mark

Q23. Which test result is criterion-referenced?

(A) "Sana scored higher than 70 percent of her class."
(B) "Sana is ranked third out of 40 students."
(C) "Sana's score is one standard deviation above the mean."
(D) "Sana correctly performed all five steps of the laboratory safety procedure."

Q24. A common criticism of using only norm-referenced tests for young learners is that they:

(A) Are too easy to construct
(B) Give too much detail about skills
(C) Encourage competition and labelling, while saying little about what a child needs to learn next
(D) Cannot be scored objectively

Q25. A grade boundary set by taking the top 20 percent of candidates, regardless of their absolute scores, reflects:

(A) A criterion-referenced standard
(B) A norm-referenced decision rule
(C) A diagnostic cut-score
(D) A formative judgement

Q26. For a mastery-learning programme, where students progress only after showing they can perform a skill, the appropriate reporting approach is mainly:

(A) Percentile ranks only
(B) Standard scores only
(C) Class ranking lists only
(D) Criterion-referenced mastery judgements

Test Construction: Table of Specifications and Item Writing

Q27. When building a table of specifications, a teacher first lists the content topics down one side and, across the top, the:

(A) The difficulty value of each test item
(B) The marks allocated to each question
(C) Cognitive levels or learning objectives to be tested for each topic
(D) The names of the textbook chapters the items come from

Q28. A test has 40 items. The specification table allocates 25 percent to Topic X. How many items should cover Topic X?

(A) 10
(B) 25
(C) 15
(D) 4

Q29. A teacher skips the specification table and writes questions from memory. The most likely result is a test that:

(A) Is perfectly balanced across topics
(B) Is too short to score
(C) Over-samples some topics and under-samples others, weakening content validity
(D) Automatically becomes norm-referenced

Q30. Which item-writing guideline helps prevent students from answering correctly by test-taking tricks rather than knowledge?

(A) Vary the position of the correct answer and keep options similar in length and style
(B) Always place the correct answer in position C
(C) Make the correct option the most detailed one
(D) Use unfamiliar vocabulary in distractors only

Q31. A stem reads: "Which of the following statements about the water cycle is true, except..." This stem is problematic mainly because it:

(A) Tests too many topics at once
(B) Combines a positive and a negative cue in a confusing way, increasing misreading
(C) Is too short
(D) Uses a multiple-choice format

Q32. An item writer should avoid absolute words such as "always" and "never" in options because such options are:

(A) More precise than other options
(B) Too difficult for students to read
(C) Used mainly in essay questions
(D) Often easily recognised as false, giving an unintended clue

Q33. The stem of a good item should contain:

(A) Two unrelated problems to save space
(B) A joke to relax students
(C) Irrelevant background detail to test reading speed
(D) The central problem clearly stated, with only the information needed to answer it

Q34. When options must be numerical, good practice is to arrange them in:

(A) Random order with no pattern
(B) Order of the teacher's preference
(C) Ascending or descending numerical order, so students compare values rather than position
(D) Order of difficulty of calculation

Q35. A teacher reviews a draft test and finds that in 18 of 20 items the correct answer is option B. The teacher should:

(A) Rebalance the key positions across options, so position does not become a clue
(B) Leave the test unchanged, since students will not notice
(C) Add more B options to the remaining items
(D) Remove option B from all items

Q36. After writing a test, the most useful review step before administration is to:

(A) Check that the total marks add up to the required total
(B) Have a colleague check items for clarity, accuracy, bias and alignment with the specification table
(C) Read through the items once for spelling errors
(D) Compare the test with last year's paper item by item

Item Types and Their Uses

Q37. A multiple-choice item that asks students to choose the best explanation for an observed result is mainly measuring:

(A) Application and understanding, not mere recall, when the situation is new to students
(B) Recall of science facts learned in class
(C) Speed and accuracy in written work
(D) Ability to follow laboratory instructions

Q38. Short-answer items are particularly useful for testing:

(A) Long, structured arguments
(B) Group discussion skills
(C) Recall of specific terms, facts or brief procedures that have a definite answer
(D) Drawing ability only

Q39. A limitation of short-answer items, compared with multiple-choice items, is that they:

(A) Cover fewer topics in the same time than multiple-choice items
(B) Require students to spell every answer correctly to gain marks
(C) Give students no chance to show partial understanding
(D) Take longer to score and allow varied wording that can make scoring less consistent

Q40. An essay question that asks students to evaluate two policies and defend a judgement is best suited to assessing:

(A) Spelling accuracy
(B) Higher-order thinking: analysis, evaluation and organisation of an argument
(C) Recall of one date
(D) Speed in completing many items

Q41. A restricted-response essay, which limits the scope and length of the answer, is preferred over an extended-response essay when the teacher wants:

(A) Maximum freedom for students to choose any topic
(B) A single overall impression mark for each script
(C) More comparable answers and more consistent scoring across students
(D) More freedom for markers to reward originality

Q42. Matching items are best used to test:

(A) Extended argumentation
(B) Laboratory manipulative skills
(C) Creative writing
(D) Associations between two sets of related facts, such as terms and definitions

Q43. A completion item ("Water boils at ___ degrees Celsius at sea level") mainly tests:

(A) Evaluation of a policy
(B) Recall of a specific fact in a brief, supplied-response form
(C) Organisation of a long argument
(D) Practical laboratory skill

Item Analysis: Difficulty, Discrimination and Distractors

Q44. An item answered correctly by only 15 percent of students, with a difficulty index of 0.15, is best described as:

(A) Very difficult, and worth checking for ambiguity, an error, or content not yet taught
(B) Ideally easy for a classroom test
(C) Automatically invalid
(D) A good discriminator regardless of its wording

Q45. If 45 out of 150 students answer an item correctly, its difficulty index is:

(A) 0.45
(B) 0.30
(C) 0.15
(D) 0.70

Q46. The top 25 students answer an item correctly 20 times and the bottom 25 answer it correctly 8 times. Its discrimination index is:

(A) 0.48
(B) 0.32
(C) 0.80
(D) 0.12

Q47. An item with a discrimination index close to zero tells the teacher that the item:

(A) Is too easy for everyone
(B) Is the best item in the test
(C) Does not separate higher from lower achievers and should be revised or removed
(D) Was scored incorrectly for all students

Q48. A distractor that attracts mainly lower-achieving students while higher achievers choose the key is best described as:

(A) A non-functioning distractor
(B) A mis-keyed option
(C) An irrelevant option
(D) A functioning, plausible distractor doing its intended job

Q49. After item analysis, a teacher would normally decide for each item to:

(A) Reuse the test unchanged, without acting on the item statistics
(B) Reject any item that proved difficult for the class, and keep the rest unchanged
(C) Decide each item on its difficulty index alone, ignoring discrimination and distractor performance
(D) Retain strong items, revise marginal items, and reject or rewrite flawed items

Q50. The main purpose of carrying out item analysis after a test is to:

(A) Rank students more harshly
(B) Improve the quality of the test for future use and identify items that misled students
(C) Set the pass mark for the next test
(D) Publish the test items for public practice

Validity, Reliability, Objectivity and Practicability

Q51. A test claims to cover the full Grade 9 biology syllabus, but its items come only from two chapters. Its main weakness is in:

(A) Content validity, because the items do not adequately represent the claimed domain
(B) Test-retest reliability, because scores would change
(C) Face validity only, because students disliked the paper
(D) Practicability, because the paper was too short

Q52. Evidence for construct validity of a test of scientific reasoning would most plausibly include that:

(A) Scores are higher for students who memorise textbook definitions
(B) Scores correlate strongly with handwriting quality
(C) Scores relate as theory predicts: higher for students with more science training, and related to, but distinct from, general reading scores
(D) Scores are the same for students with different amounts of science training

Q53. A teacher-training admission test is checked against trainees' teaching-practice grades a year later. A strong relationship would support its:

(A) Predictive validity
(B) Content validity
(C) Face validity
(D) Split-half reliability

Q54. A new quick screening test is validated against a full diagnostic assessment given in the same week. This comparison provides evidence of:

(A) Test-retest reliability
(B) Content validity
(C) Face validity
(D) Concurrent validity

Q55. Face validity matters in practice mainly because a test that does not look relevant to test takers may:

(A) Reduce the number of items in the test
(B) Reduce motivation and cooperation, even if its technical quality is sound
(C) Make the test easier for weaker students
(D) Increase the time allowed for the test

Q56. A weakness of test-retest reliability is that, if the interval between administrations is very short, scores may be inflated by:

(A) Changes in the syllabus
(B) A different examiner
(C) Practice and memory effects from the first administration
(D) A larger norm group

Q57. To estimate split-half reliability, a teacher correlates scores on odd-numbered items with scores on even-numbered items, then applies:

(A) The difficulty index formula
(B) A percentile rank conversion
(C) The Spearman-Brown formula, to estimate reliability for the full test length
(D) A standard deviation only

Q58. KR-21 differs from KR-20 in that KR-21:

(A) Is used only for essay tests
(B) Assumes all items are of equal difficulty, giving a quicker but less exact estimate for dichotomous items
(C) Requires no item data at all
(D) Measures validity rather than reliability

Q59. Objectivity in testing is best shown when:

(A) The test covers the full syllabus fairly
(B) The test is administered under the same conditions for all pupils
(C) The test items are all of the same difficulty
(D) Scoring does not depend on who marks the test, because the key or rubric leaves little room for personal judgement

Q60. A school chooses a short, inexpensive test that can be given in one period, even though a longer test would be slightly more reliable. This choice prioritises practicability, and it is acceptable only if:

(A) The chosen test remains sufficiently valid and reliable for its intended decision, since a valid test must also be reliable
(B) The test is made free of charge
(C) The test is printed clearly and easy to administer
(D) The test covers fewer topics than the longer test

Answer Key

Q1: C	Q11: C	Q21: A	Q31: B	Q41: C	Q51: A
Q2: A	Q12: B	Q22: B	Q32: D	Q42: D	Q52: C
Q3: D	Q13: B	Q23: D	Q33: D	Q43: B	Q53: A

Q4: B	Q14: C	Q24: C	Q34: C	Q44: A	Q54: D
Q5: D	Q15: D	Q25: B	Q35: A	Q45: B	Q55: B
Q6: B	Q16: A	Q26: D	Q36: B	Q46: A	Q56: C
Q7: A	Q17: D	Q27: C	Q37: A	Q47: C	Q57: C
Q8: C	Q18: A	Q28: A	Q38: C	Q48: D	Q58: B
Q9: A	Q19: B	Q29: C	Q39: D	Q49: D	Q59: D
Q10: D	Q20: C	Q30: A	Q40: B	Q50: B	Q60: A

Worked explanations for all 60 questions:
quaidianpoint.com/pedagogy-curriculum-development-assessment-practice-test-8/