

Educational Measurement & Evaluation

Practice Test 3

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

Measurement, Assessment and Evaluation

Q1. A headteacher studies attendance, pass rates and parent feedback, then decides whether to extend a remedial programme. This decision process is best called:

- (A) Scoring
- (B) Item analysis
- (C) Ranking
- (D) Evaluation

Q2. A teacher asks short oral questions during a lesson to check understanding before moving on. This is:

- (A) Final examination
- (B) Formative assessment
- (C) Certification
- (D) External moderation

Q3. Which example is measurement?

- (A) Judging a school as excellent
- (B) Recommending promotion
- (C) Recording 27 correct answers out of 30
- (D) Deciding a course should close

Q4. A student meets the fixed standard of 60 percent for a skills certificate. This is an example of:

- (A) Criterion-referenced judgement
- (B) Peer ranking
- (C) Norm-referenced judgement
- (D) Self-grading only

Q5. A teacher keeps a diary of observations about a struggling reader across six weeks. This record supports:

- (A) One-off summative grading only
- (B) Ongoing assessment of progress
- (C) Standardised ranking
- (D) Item banking

Q6. Which use of a test is summative?

- (A) Adjusting tomorrow lesson plan
- (B) Giving feedback on a draft
- (C) Diagnosing a misconception mid-unit
- (D) Awarding final grades at term end

Q7. A new student is tested to decide which class level suits her best. This test is used for:

- (A) Certification
- (B) Final grading
- (C) Placement
- (D) Programme closure

Q8. Assessment differs from a single test score because assessment may include:

- (A) Observations, projects and discussions as well as tests
- (B) Only multiple-choice items
- (C) Only final examinations
- (D) Only numerical marks

Q9. A teacher compares Faraz score with the scores of all students who took the same national test. This is:

- (A) Criterion referencing
- (B) Norm referencing
- (C) Diagnostic labelling
- (D) Formative feedback

Q10. Which statement best describes diagnostic assessment?

- (A) It certifies course completion
- (B) It ranks whole schools
- (C) It identifies specific strengths, gaps and causes of difficulty
- (D) It replaces all teaching

Q11. A rubric used to judge project quality against described levels is mainly a tool for:

- (A) Criterion-referenced assessment
- (B) Random ranking
- (C) Attendance recording
- (D) Item difficulty analysis

Q12. The main value of continuous assessment is that it:

- (A) Depends on one examination day
- (B) Ignores classroom work
- (C) Removes feedback entirely
- (D) Reduces reliance on a single high-stakes test

Scales of Measurement

Q13. Classifying textbooks by subject - maths, science, English - uses which scale?

- (A) Nominal scale
- (B) Ordinal scale
- (C) Interval scale
- (D) Ratio scale

- Q14. A league table listing schools from first to tenth, with no score gaps shown, is:**
- (A) Nominal data
 - (B) Ratio data
 - (C) Ordinal data
 - (D) Interval data
- Q15. Scores on an IQ-type scale, where equal point gaps are assumed but zero is arbitrary, are treated as:**
- (A) Nominal
 - (B) Ordinal
 - (C) Categorical only
 - (D) Interval
- Q16. The number of minutes a student spends reading each day is measured on a:**
- (A) Nominal scale
 - (B) Ratio scale
 - (C) Ordinal scale
 - (D) Interval scale only
- Q17. Which statement about a ratio scale is true?**
- (A) Categories cannot be ordered
 - (B) Equal intervals are absent
 - (C) Zero is always arbitrary
 - (D) A value of zero means absence of the quantity
- Q18. For nominal data, the only central tendency that is always meaningful is the:**
- (A) Mean
 - (B) Mode
 - (C) Median
 - (D) Standard deviation
- Q19. A teacher converts letter grades to ranks for a prize list. The resulting data are:**
- (A) Ordinal
 - (B) Ratio
 - (C) Interval
 - (D) Nominal
- Q20. On an interval scale, the difference between 30 and 40 is:**
- (A) Twice the difference between 10 and 15
 - (B) Meaningless
 - (C) The same as the difference between 70 and 80
 - (D) Half the difference between 20 and 40
- Q21. Which variable is ordinal?**
- (A) Number of pages in a book
 - (B) Weight in grams
 - (C) Level of agreement from strongly disagree to strongly agree
 - (D) Eye colour categories

Q22. A researcher can say one school library has twice as many books as another because book counts are on a:

- (A) Nominal scale
- (B) Ordinal scale
- (C) Interval scale
- (D) Ratio scale

Q23. Which statistic should NOT be used with ordinal ranks alone?

- (A) Mean, because it assumes equal intervals
- (B) Median
- (C) Mode
- (D) Frequency count

Q24. Temperature in Kelvin, unlike Celsius, is a ratio scale because:

- (A) It uses larger numbers
- (B) Its zero is absolute zero, meaning no thermal energy
- (C) It is measured with thermometers
- (D) It has no equal intervals

Validity

Q25. A test of English writing that is scored only on handwriting neatness has a validity problem because it measures:

- (A) Content coverage
- (B) An irrelevant trait instead of writing skill
- (C) Test length
- (D) Scoring speed

Q26. Content validity is best improved by:

- (A) Careful sampling of the whole syllabus when writing items
- (B) Making items harder
- (C) Shortening the test sharply
- (D) Using only true-false items

Q27. A new aptitude test is validated against supervisor ratings collected in the same month. This is:

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

Q28. A scholarship test is judged by how well it forecasts university success two years later. This is:

- (A) Face validity
- (B) Content validity
- (C) Predictive validity
- (D) Inter-rater reliability

Q29. A science reasoning test correlates with lab performance but not with spelling skill. This pattern best supports:

- (A) Construct validity
- (B) Scoring reliability only
- (C) Face validity only
- (D) Item difficulty

Q30. Which factor can weaken predictive validity?

- (A) Clear test directions
- (B) A reliable test
- (C) A relevant criterion
- (D) Changes in teaching or support between test and criterion

Q31. A test appears professional and relevant to candidates, but no other validity evidence exists. It has only:

- (A) Content validity
- (B) Face validity
- (C) Predictive validity
- (D) Construct validity

Q32. Validity is best thought of as a property of:

- (A) The test booklet alone
- (B) The printing quality
- (C) The interpretation and use of scores
- (D) The number of items alone

Q33. A test that is reliable but invalid for a given purpose will:

- (A) Give random scores
- (B) Be easy to use
- (C) Cover the syllabus fully
- (D) Give consistent but misleading results for that purpose

Q34. Which validity question asks whether the test covers the intended domain fairly?

- (A) Does it predict future success?
- (B) Does it look suitable?
- (C) Does it represent the content domain?
- (D) Do scorers agree?

Q35. To establish construct validity, researchers often use:

- (A) Only one easy item
- (B) Factor analysis and group comparisons based on theory
- (C) Counting correct answers only
- (D) Timing the test only

Q36. A criterion that itself is unreliable will usually make validity appear:

- (A) Lower than the true validity
- (B) Higher than the true validity
- (C) Unaffected
- (D) Perfect

Reliability

Q37. Reliability is necessary but not sufficient for validity because:

- (A) Consistency removes all error
- (B) A valid test is always short
- (C) A test can be consistent yet measure the wrong trait
- (D) Validity needs no consistency

Q38. Two forms of a test with matched content and difficulty are used to estimate:

- (A) Face validity
- (B) Content validity
- (C) Inter-rater agreement
- (D) Parallel-forms reliability

Q39. A single test is split into two halves by odd and even numbers. This estimates:

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

Q40. KR-20 is suitable when items are scored:

- (A) By partial credit only
- (B) As correct or incorrect
- (C) By judges on a rubric
- (D) By time taken only

Q41. A test-retest study uses a gap of one day for a memory-heavy test. The main risk is:

- (A) Practice and memory effects inflating stability
- (B) Complete change in the trait
- (C) Loss of all items
- (D) Change in test length

Q42. A teacher wants to know if essay marks would be similar if another teacher marked them. She should check:

- (A) Split-half reliability
- (B) Test-retest reliability
- (C) Inter-rater reliability
- (D) Parallel-forms reliability

Q43. Adding many poor-quality items to a test will most likely:

- (A) Raise validity sharply
- (B) Improve content coverage fully
- (C) Reduce testing time
- (D) Fail to improve reliability and may harm it

Q44. A score of 65 with SEM 3 suggests the true score probably lies around:

- (A) Exactly 65
- (B) 62 to 68
- (C) 50 to 80
- (D) 0 to 100

Q45. Which change lowers the standard error of measurement?

- (A) Lower reliability
- (B) Larger score spread only
- (C) More guessing
- (D) Higher reliability

Q46. Guessing on multiple-choice items mainly adds:

- (A) True score variance
- (B) Random error variance
- (C) Content validity
- (D) Item difficulty control

Q47. A reliability coefficient of 0.95 for a short classroom quiz should be viewed with:

- (A) Full trust at once
- (B) Ignoring the result
- (C) Caution, checking method and sample before trusting it
- (D) Rejection of all quizzes

Q48. Standardising directions, timing and scoring mainly improves:

- (A) Reliability by reducing irrelevant variation
- (B) Item difficulty alone
- (C) Test length
- (D) Face appeal only

Item Analysis

Q49. If 60 of 200 students answer an item correctly, its difficulty index is:

- (A) 0.30
- (B) 0.60
- (C) 0.20
- (D) 0.40

Q50. A p value of 0.15 indicates an item that is:

- (A) Very easy
- (B) Very difficult
- (C) Moderate
- (D) Perfectly discriminating

Q51. Upper group 27 of 30 correct and lower group 9 of 30 correct gives discrimination of:

- (A) 0.30
- (B) 0.90
- (C) 0.20
- (D) 0.60

Q52. A discrimination index of 0.10 is best described as:

- (A) Excellent
- (B) Good
- (C) Weak
- (D) Negative

Q53. An item with negative discrimination should first be:

- (A) Retained unchanged
- (B) Given double marks
- (C) Placed at the start
- (D) Checked for a wrong key or ambiguous wording

Q54. Which item is most suitable for a mastery test where most students should pass?

- (A) An item with $p = 0.85$ and clear wording
- (B) An item with $p = 0.10$
- (C) An item with negative discrimination
- (D) An item with a dead distractor

Q55. Distractor analysis that shows one option chosen by 40 percent of low scorers and 5 percent of high scorers indicates:

- (A) A dead distractor
- (B) A faulty key
- (C) A functioning, plausible distractor
- (D) An item with no discrimination

Q56. The best action for a distractor chosen by almost no one is to:

- (A) Keep it forever
- (B) Replace it with a more plausible option
- (C) Make it the key
- (D) Remove the item stem

Q57. A test with many very easy items will most likely show:

- (A) Strong discrimination throughout
- (B) Low discrimination, because most students score alike
- (C) Higher validity automatically
- (D) No need for item analysis

Q58. Item analysis is normally carried out:

- (A) Before any student takes the test
- (B) Only for essay tests
- (C) Never, if items look good
- (D) After a trial administration, before final use

Q59. An item with $p = 0.48$, $D = 0.42$ and all distractors chosen by some students is:

- (A) Too easy for any use
- (B) A flawed item to delete at once
- (C) A sound item to retain
- (D) Valid only for essays

Q60. If high scorers choose a distractor more often than the key, the test developer should suspect:

- (A) A mis-keyed or ambiguous item
- (B) An item that is too easy
- (C) A test that is too short
- (D) A group that is too large

Answer Key

Q1: D	Q11: A	Q21: C	Q31: B	Q41: A	Q51: D
Q2: B	Q12: D	Q22: D	Q32: C	Q42: C	Q52: C
Q3: C	Q13: A	Q23: A	Q33: D	Q43: D	Q53: D
Q4: A	Q14: C	Q24: B	Q34: C	Q44: B	Q54: A
Q5: B	Q15: D	Q25: B	Q35: B	Q45: D	Q55: C
Q6: D	Q16: B	Q26: A	Q36: A	Q46: B	Q56: B
Q7: C	Q17: D	Q27: D	Q37: C	Q47: C	Q57: B
Q8: A	Q18: B	Q28: C	Q38: D	Q48: A	Q58: D
Q9: B	Q19: A	Q29: A	Q39: A	Q49: A	Q59: C
Q10: C	Q20: C	Q30: D	Q40: B	Q50: B	Q60: A

Worked explanations for all 60 questions:

quaidianpoint.com/pedagogy-educational-measurement-evaluation-practice-test-3/