

Educational Measurement & Evaluation

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

Norm-Referenced and Criterion-Referenced Testing

Q1. A norm-referenced score is best interpreted as:

- (A) The number of objectives the pupil has mastered
- (B) The percentage of the syllabus the pupil knows
- (C) The pupil's exact marks out of the total
- (D) The pupil's standing relative to the norm group

Q2. Which situation calls for criterion-referenced testing rather than norm testing?

- (A) Choosing the best student in the district for a prize
- (B) Ranking schools for a newspaper list
- (C) Selecting one captain from a large team
- (D) Deciding whether a student is ready to move on after mastering a unit

Q3. A key advantage of criterion-referenced testing is that it

- (A) Gives clear information about what a student can actually do
- (B) Always produces a wide spread of scores
- (C) Forces a fixed share of students to fail
- (D) Hides gaps in learning behind a rank

Q4. A key advantage of norm-referenced testing is that it

- (A) Proves mastery of every taught objective
- (B) Removes the need for any norm group
- (C) Allows comparison when a common standard is hard to fix
- (D) Works without any scoring key

Q5. A norm-referenced test can mislead a parent if the parent reads a percentile as

- (A) A rank among peers
- (B) A percentage of items answered correctly
- (C) A position in the norm group
- (D) A comparison score

Q6. A criterion-referenced test can mislead if the standard is

- (A) Set arbitrarily without a sound basis
- (B) Shared with students in advance
- (C) Linked to clear objectives
- (D) Applied equally to all students

Q7. Can the same test be interpreted in both ways?

- (A) No, a test can never support two interpretations
- (B) Only if the test has no scoring key
- (C) Yes, if both a norm group and a criterion are available for its scores
- (D) Only if the test is oral

Q8. In a class where every student masters the objectives, a norm-referenced test will still

- (A) Rank students against each other
- (B) Report that all have failed
- (C) Become invalid for all use
- (D) Show identical scores for everyone

Q9. A teacher who wants both ranking and mastery information should use

- (A) Only a class rank list
- (B) Only a total mark
- (C) Norm scores plus a criterion report for objectives
- (D) Only attendance records

Grading Systems and GPA

Q10. Which grading approach best supports a mastery-learning classroom?

- (A) Forced ranking with a fixed number of top grades
- (B) Grading by attendance only
- (C) A single high-stakes test with no feedback
- (D) Criterion-based grading with reassessment opportunities

Q11. A disadvantage of relative grading is that it

- (A) Makes standards clear to every student
- (B) Always raises overall achievement
- (C) Can discourage students from helping peers
- (D) Removes competition completely

Q12. A disadvantage of absolute grading is that it

- (A) Always ranks students unfairly
- (B) May be hard to set fair cut-offs for a new or complex course
- (C) Requires a very large norm group
- (D) Depends on the performance of other classes

Q13. A student argues that a B in a very hard course should count more than a B in an easy course. This argument supports

- (A) Pass-fail grading only
- (B) Weighted grading
- (C) Ranking by age
- (D) Grading by attendance

Q14. Pass-fail grading is most suitable when the main goal is

- (A) Ranking students for scholarships
- (B) Comparing students across schools precisely
- (C) Judging whether a minimum standard is met, not fine ranking
- (D) Awarding honours distinctions

Q15. A teacher combines test scores, projects and classwork into a final grade. This practice improves grading because it

- (A) Makes grading faster to explain
- (B) Samples learning more broadly than one test alone
- (C) Guarantees every student an A
- (D) Removes the need for a syllabus

Q16. If two teachers grade the same essay very differently, the grading lacks

- (A) A grade scale
- (B) A course title
- (C) Reliability
- (D) A pass mark

Q17. A grade that reflects behaviour and neatness more than learning lacks

- (A) A letter symbol
- (B) A numerical value
- (C) Validity as a measure of achievement
- (D) A place on the transcript

Q18. A school changes from letter grades to detailed written reports. A likely benefit is

- (A) Faster ranking of all students
- (B) Richer information about strengths and next steps
- (C) A simpler GPA calculation
- (D) Less work for teachers

Standardized Tests and Their Characteristics

Q19. Standardized tests help policymakers mainly by

- (A) Providing comparable data across many schools
- (B) Replacing teacher judgement in every lesson
- (C) Measuring school building quality
- (D) Counting student enrolment

Q20. A standardized test with poor validity for a local curriculum should be used

- (A) Cautiously, with other evidence, or not at all for that purpose
- (B) As the only measure for promotion
- (C) Without any caution
- (D) To replace the local syllabus

Q21. Test bias is a serious concern because a biased item may

- (A) Raise scores for every student equally
- (B) Make the test shorter
- (C) Lower scores for a group for reasons unrelated to the skill tested
- (D) Improve the test for all groups

Q22. Standardization does NOT guarantee that a test is

- (A) Scored in the same way
- (B) Valid for every purpose a user might choose
- (C) Given under similar conditions
- (D) Interpreted with norms

Q23. A school that prepares students only by drilling past standardized items risks

- (A) Improving all learning automatically
- (B) Teaching to the test instead of the wider curriculum
- (C) Making tests more valid
- (D) Removing test anxiety fully

Q24. Norms become outdated when

- (A) More students take the test
- (B) The test is printed in colour
- (C) The population or curriculum changes over time
- (D) Teachers read the manual

Q25. A standardized achievement test differs from a teacher-made test mainly in that it

- (A) Is always shorter
- (B) Is never used for grading
- (C) Contains no multiple-choice items
- (D) Is developed, tried out and normed on a large sample

Q26. For a high-stakes decision, a single standardized score should be

- (A) Accepted without question
- (B) Ignored completely
- (C) Checked against other evidence before deciding
- (D) Used only if it is low

Formative, Summative and Diagnostic Evaluation

Q27. A teacher gives a quiz, finds a common error, and reteaches the topic the same week. The quiz functioned as

- (A) Formative assessment
- (B) A final certification test
- (C) An admission test
- (D) A school census

Q28. A final project graded at the end of term and reported to parents functioned as

- (A) Daily formative feedback
- (B) Summative assessment
- (C) A diagnostic interview
- (D) A placement screening

Q29. A counsellor tests a struggling reader's phonics, vocabulary and fluency separately. This is mainly

- (A) Summative grading
- (B) Diagnostic assessment
- (C) Norm-referenced ranking
- (D) A final board result

Q30. Which statement best separates formative from summative evaluation?

- (A) Formative is always oral; summative is always written
- (B) Formative uses grades; summative never uses grades
- (C) Formative happens once a year; summative happens daily
- (D) Formative guides ongoing learning; summative judges completed learning

Q31. A diagnostic test that only gives a total score without sub-skill detail is weak because it

- (A) Does not show where the difficulty lies
- (B) Is too easy for all students
- (C) Cannot be scored quickly
- (D) Uses too many items

Q32. Peer feedback during group work is formative when students

- (A) Keep it secret from the teacher
- (B) Use it to improve their work before submission
- (C) Use it to rank group members
- (D) Ignore it completely

Q33. A school uses baseline testing at entry to group students for support. This use is closest to

- (A) Placement or diagnostic use before teaching
- (B) Summative certification
- (C) Final grading
- (D) A public ranking

Q34. Summative assessment can still help future learning if

- (A) Scores are filed and never discussed
- (B) Only the top students see their results
- (C) It is given without any instructions
- (D) Results are reviewed to improve the next course or unit

Q35. Which practice mixes the types poorly?

- (A) Using quizzes to guide reteaching
- (B) Using a diagnostic check before remediation
- (C) Using a final test to certify a course
- (D) Using a single end-of-year score alone to plan daily remedial help

Test Construction Steps

Q36. A test constructor who skips the blueprint risks building a test that

- (A) Is automatically more reliable
- (B) Has perfect content coverage
- (C) Over-tests some topics and ignores others
- (D) Needs no tryout

Q37. Writing items before defining outcomes risks building a test that

- (A) Is automatically valid
- (B) Requires no scoring key
- (C) Measures trivia instead of intended learning
- (D) Cannot be printed

Q38. A multiple-choice stem that asks two questions at once is faulty because

- (A) Students may know one part and not the other
- (B) It is always too short
- (C) It cannot have four options
- (D) It is always too easy

Q39. An item whose correct option is much longer than the distractors is faulty because

- (A) Long answers are always wrong
- (B) Test-wise students may pick it by length alone
- (C) Short options are always correct
- (D) Length has no effect on choices

Q40. A tryout sample should resemble

- (A) Only the highest achievers
- (B) Only students from one school
- (C) The students who will take the final test
- (D) Adults who no longer study the subject

Q41. Item analysis after tryout helps a constructor to

- (A) Increase the pass mark secretly
- (B) Change the syllabus after testing
- (C) Award grades without scoring
- (D) Keep, revise or drop items based on evidence

Q42. A test that is too long for the time allowed will mainly harm

- (A) The difficulty of individual items alone
- (B) The total number of pages alone
- (C) Validity, because unfinished items do not measure learning
- (D) The speed of scoring alone

Q43. Assembling the final test requires attention to

- (A) Item order, layout, instructions and time allowed
- (B) Only the total number of pages
- (C) Only the name of the school
- (D) Only the date of printing

Q44. A field test that shows a key error in one item should lead to

- (A) Keeping the error to test attention
- (B) Removing the whole topic from teaching
- (C) Giving all students full marks for the paper
- (D) Correcting the key before final use

Basic Statistics for Teachers

Q45. A teacher finds the mean is much higher than the median. This suggests the data are

- (A) Perfectly symmetrical
- (B) Pulled upward by some high extreme scores
- (C) All identical
- (D) Impossible to summarise

Q46. A class has scores 60, 62, 64, 66, 68. The standard deviation is

- (A) Very large, because scores increase
- (B) Zero, because scores differ
- (C) Impossible to estimate
- (D) Small, because scores cluster tightly

Q47. If a score has a z-score of +1.5, it lies

- (A) One and a half standard deviations above the mean
- (B) One and a half marks above the mean
- (C) At the 150th percentile
- (D) Fifteen points above the highest score

Q48. A percentile rank of 25 and a percent score of 25 are

- (A) Different: rank is position, percent is proportion correct
- (B) The same in all cases
- (C) Both equal to 25 marks
- (D) Both impossible on one test

Q49. In a normal distribution, the mean, median and mode are

- (A) Far apart at the extremes
- (B) Approximately equal and at the centre
- (C) Equal to the range
- (D) Always zero

Q50. A skewed distribution with a long tail to the right has a mean that is

- (A) Less than the median
- (B) Equal to the mode
- (C) Unrelated to the median
- (D) Greater than the median

Q51. The interquartile range measures the spread of

- (A) The middle 50 percent of scores
- (B) All scores from lowest to highest
- (C) Only the top quarter of scores
- (D) Only scores above the mean

Q52. A teacher wants a single value for typical performance in a class with one extreme outlier. The best choice is the

- (A) Mean
- (B) Range
- (C) Highest score
- (D) Median

Q53. A frequency distribution is most useful for showing

- (A) Only the name of the top student
- (B) How scores are spread across values or intervals
- (C) Only the pass mark
- (D) The date of the test

Q54. Two classes have the same mean but different standard deviations. The class with the larger deviation has scores that are

- (A) More clustered
- (B) Higher on average
- (C) Lower on average
- (D) More spread out

Evaluation Models

Q55. Tyler's model works best when objectives are

- (A) Kept secret from teachers
- (B) Clear, measurable and agreed in advance
- (C) Changed after results are known
- (D) Written only after teaching ends

Q56. A weakness of Tyler's model for a new programme is that it

- (A) Requires no objectives at all
- (B) Forbids the use of tests
- (C) Gives little guidance on needs assessment before objectives are set
- (D) Works only for mathematics

Q57. CIPP is called a decision-oriented model because its evaluations are meant to

- (A) Support decisions at planning, implementation and outcome stages
- (B) Produce a single final grade
- (C) Rank students for prizes
- (D) Replace all classroom tests

Q58. In CIPP, a decision to continue, change or stop a programme mainly uses

- (A) Only attendance data
- (B) Only textbook prices
- (C) Only staff opinions
- (D) Product evaluation evidence

Q59. Which of the following is an example of input evaluation?

- (A) Judging final examination results after a programme ends
- (B) Checking daily attendance during a programme
- (C) Asking parents how they felt after a programme ended
- (D) Comparing two reading schemes' costs and materials before adopting one

Q60. Process evaluation helps decision makers to

- (A) Detect implementation problems early and adjust the programme
- (B) Set admission standards for next year
- (C) Judge only final results
- (D) Choose a school building

Answer Key

Q1: D	Q11: C	Q21: C	Q31: A	Q41: D	Q51: A
Q2: D	Q12: B	Q22: B	Q32: B	Q42: C	Q52: D

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

Worked explanations for all 60 questions:

quaidianpoint.com/pedagogy-educational-measurement-evaluation-practice-test-8/