Psychological Assessment Study Guide

20 Essay-Style Prompts & Answers • 20 MCQs with Explanations • Reliability • Validity • Norms • IRT • Ethics

How to use this guide: Start with the essay-style prompts to solidify core assessment concepts (reliability, validity, norms, score interpretation, fairness). Then run the MCQs under timed conditions (~60–90 s each). Use explanations to spot patterns—standard errors, cut-scores, item analysis, ROC metrics, and construct validity evidence. For full-length exams with detailed rationales and printable sheets, upgrade below.

Essay-Style Prompts & Answers (20)

1. Reliability Overview

Prompt: Define and differentiate test-retest, interrater, and internal-consistency reliability; give one use case for each.
Answer (concise): Test-retest: stability across time (e.g., trait measures). Interrater: agreement between raters (e.g., scoring open-ended responses). Internal consistency: item intercorrelation (e.g., Likert scales).

2. Cronbach’s Alpha & Spearman-Brown

Prompt: Explain what alpha estimates and when Spearman-Brown correction is applied.
Answer (concise): Alpha estimates lower-bound internal consistency assuming tau-equivalence. Spearman-Brown predicts reliability when test length changes (e.g., split-half correction or planned item additions).

3. Standard Error of Measurement (SEM)

Prompt: Define SEM and show how to compute a 95% confidence interval around an observed score.
Answer (concise): SEM = SD × √(1−r). 95% CI ≈ X ± 1.96×SEM, yielding the likely range for true scores.

4. Validity Types

Prompt: Contrast content, criterion (predictive/concurrent), and construct validity.
Answer (concise): Content: coverage of domain; Criterion: correlation with outcomes (predictive/concurrent); Construct: theory-consistent relations (convergent/discriminant, factor structure, known-groups).

5. Incremental Validity

Prompt: Define and give an example using hierarchical regression.
Answer (concise): Incremental validity is added predictive value beyond existing measures; e.g., a new executive-function scale adds R^2 over IQ and education when predicting job performance.

6. Sensitivity/Specificity & Predictive Values

Prompt: Explain sensitivity/specificity vs. PPV/NPV and the role of base rates.
Answer (concise): Sensitivity/specificity are test properties; PPV/NPV depend on prevalence—low base rates lower PPV even with high sensitivity/specificity.

7. Cut Scores

Prompt: Describe Angoff method basics and a risk of misclassification.
Answer (concise): Experts estimate probability a minimally qualified candidate answers each item correctly; averaged to set cut. Risk: standard-setting bias and error around the cut leading to false positives/negatives.

8. Standardization & Norms

Prompt: Differentiate norm-referenced from criterion-referenced interpretation; list common norms.
Answer (concise): Norm-referenced compares to a reference group (percentiles, z, T-scores, stanines); criterion-referenced compares to fixed performance standards.

9. Scaling & Score Types

Prompt: Explain z, T, and stanines and why linear transformations preserve rank.
Answer (concise): $z = (X-M)/SD$; $T = 50 + 10z$; stanines = 1–9 scale. Linear transforms preserve order and relative distances; they don’t change norm position.

10. Item Difficulty and Discrimination

Prompt: Define p-value and discrimination index; what are desirable ranges in CTT?
Answer (concise): p-value = proportion correct (difficulty). Discrimination = item-total correlation; desirable p around .3–.8 with positive discrimination (>.2 typical).

11. Classical Test Theory vs. IRT

Prompt: State one advantage of IRT over CTT for computerized adaptive testing (CAT).
Answer (concise): IRT models item-level parameters (difficulty b, discrimination a, sometimes guessing c), enabling tailored item selection and ability estimates with known precision.

12. Factor Analysis Basics

Prompt: Differentiate exploratory (EFA) and confirmatory (CFA) factor analysis.
Answer (concise): EFA discovers latent structure without a priori constraints; CFA tests a hypothesized structure with fit indices and cross-loadings constrained.

13. Bias, Fairness, DIF

Prompt: Define differential item functioning (DIF) and how it's detected.
Answer (concise): DIF occurs when groups with same ability have different item probabilities. Detected by IRT-based methods or Mantel–Haenszel/ logistic regression, followed by content review.

14. Response Styles & Faking

Prompt: Name two response biases and one mitigation strategy.
Answer (concise): Acquiescence and social desirability; mitigate with balanced keyed items, validity scales, warnings, and forced-choice formats.

15. Ethics & APA Standards

Prompt: List key ethical principles relevant to testing and one example each.
Answer (concise): Competence (appropriate training), Informed consent (purpose/uses), Privacy (secure data), Fairness (accommodations), Test security (protect items).

16. Administration & Accommodations

Prompt: Give two examples of valid accommodations that preserve construct.
Answer (concise): Extended time for processing-speed deficits; large-print/reader for visual impairments—if the construct isn't speed/vision dependent.

17. Interpretation & Feedback

Prompt: Outline best practices for feedback to clients/stakeholders.
Answer (concise): Explain purpose and limits, present scores with CIs, focus on strengths and needs, avoid over-pathologizing, use plain language, invite questions.

18. Base Rates & Clinical Utility

Prompt: Why do low base rates complicate decision-making?
Answer (concise): With rare conditions, false positives dominate even with good tests; decision thresholds and multiple evidence sources are needed.

19. Test Security & Retesting

Prompt: Discuss retest intervals and item exposure concerns.
Answer (concise): Short retests inflate practice effects; rotate parallel forms and enforce security to avoid item harvesting and compromised validity.

20. Cross-Cultural Assessment

Prompt: Name two steps to improve cultural fairness in testing.
Answer (concise): Use translated/adapted instruments with equivalence studies; include local norms and qualitative context to interpret results.

Multiple-Choice Questions (20) — with Explanations

1. A test yields similar scores when administered twice, two weeks apart. This best reflects…
A. Internal consistency
B. Interrater reliability
C. Test-retest reliability
D. Content validity
Answer: C
Explanation: Stability across time is test-retest reliability.

2. Cronbach’s alpha primarily indexes…
A. Temporal stability
B. Agreement among raters
C. Internal consistency among items
D. Predictive validity
Answer: C
Explanation: Alpha estimates internal consistency (average inter-item covariance).

3. A 95% CI for a score uses which quantity?
A. SEM
B. SD of the norm group only
C. Interquartile range
D. Item difficulty
Answer: A
Explanation: CI around observed score uses the Standard Error of Measurement.

4. Which evidence supports content validity most directly?
A. Correlation with a future criterion
B. Expert judgment of domain coverage
C. High alpha coefficient
D. Test-retest correlation
Answer: B
Explanation: Subject-matter experts evaluate content coverage and relevance.

5. If a cut score is set too high, which error increases?
A. False positives
B. False negatives
C. True positives
D. Sensitivity
Answer: B
Explanation: Raising the cut risks missing qualified candidates (false negatives).

6. In a low-prevalence setting, which is most likely to decrease?
A. Specificity
B. Sensitivity
C. Positive predictive value
D. Negative predictive value
Answer: C
Explanation: PPV drops as base rate decreases even with same sensitivity/specificity.

7. Norm-referenced interpretation means…
A. Comparing scores to a fixed standard
B. Comparing to a reference group distribution
C. Mastery/no-mastery only
D. Criterion alignment
Answer: B
Explanation: Norm-referenced compares an individual to group norms.

8. Transforming z-scores to T-scores primarily…
A. Changes rank order
B. Reduces measurement error
C. Re-scales to mean 50 SD 10 without changing rank
D. Improves validity
Answer: C
Explanation: Linear scaling preserves rank and relative distances.

9. An item with p = .95 likely has…
A. Good discrimination
B. Appropriate difficulty
C. Too easy; low discrimination
D. Too hard; low discrimination
Answer: C
Explanation: Very high p means most got it correct—often low discrimination.

10. A major advantage of IRT for CAT is…
A. Fixed test length for all
B. Ability estimates independent of item set given calibrated bank
C. No need for calibration
D. Eliminates SEM
Answer: B
Explanation: IRT allows tailored items and ability estimates tied to item parameters.

11. EFA is preferred when…
A. You have a strong a priori model
B. You need to explore latent structure without constraints
C. You want to test exact factor loadings
D. You estimate path models among factors only
Answer: B
Explanation: EFA explores structure; CFA tests specified structure.

12. DIF indicates that an item…
A. Is too difficult for everyone
B. Functions differently across groups at equal ability
C. Has low reliability
D. Predicts criterion weakly
Answer: B
Explanation: DIF = bias where groups differ at the same trait level.

13. A candidate inflates socially desirable responses. Which strategy helps?
A. Use more agree-keyed items
B. Add a social desirability scale or forced-choice format
C. Remove reverse-keyed items
D. Ignore and average responses
Answer: B
Explanation: Validity scales/forced-choice reduce faking and acquiescence effects.

14. Which is the best initial step when giving feedback on a low score?
A. State the number and end session
B. Explain score meaning, include CI, and invite questions
C. Recommend therapy
D. Compare to your own experience
Answer: B
Explanation: Use plain language, context, and confidence intervals; allow questions.

15. A test for speeded processing should NOT grant which accommodation?
A. Large print
B. Alternative font for dyslexia
C. Extra time
D. Reader for visually impaired content
Answer: C
Explanation: If speed is the construct, extra time changes the construct measured.

16. Base rate neglect leads to…
A. Overreliance on prevalence
B. Underweighting prior probability in interpreting a positive test
C. Better PPV
D. Improved sensitivity
Answer: B
Explanation: Ignoring base rates inflates belief a positive means the condition is present.

17. To reduce practice effects on retest, use…
A. Identical items
B. Parallel forms
C. Shorter intervals
D. Unproctored settings
Answer: B
Explanation: Parallel/alternate forms help reduce practice effects.

18. Interrater reliability is best increased by…
A. Adding more items
B. Detailed rubrics and rater training
C. Shorter test time
D. Using only experts
Answer: B
Explanation: Clear scoring criteria and training improve agreement.

19. An ROC curve helps select a cut by trading off…
A. Reliability and validity
B. Sensitivity and specificity
C. Alpha and beta
D. Mean and SD
Answer: B
Explanation: ROC visualizes sensitivity–specificity trade-offs.

20. Construct validity is best supported by…
A. One high alpha value
B. High test-retest only
C. Convergent and discriminant correlations consistent with theory
D. Attractive test materials
Answer: C
Explanation: Construct validity draws on theory-consistent patterns of relations.