

10. MM.1 · Psychometrics and test construction

Starter Quiz ·

GRE Subject Test · GRE Psychology Name: _____ Score: _____

1. Under the parallel-halves assumptions, a test’s two comparable halves correlate 0.70. What full-test reliability does Spearman–Brown estimate, to two decimal places?

2. On one score scale, $SD = 10$ and the relevant classical reliability estimate is 0.84. Using $SEM = SD \times \sqrt{1 - r}$, what is SEM in score points?

3. What T-score corresponds to $z = -1.5$?

4. An item is answered correctly less often by a higher-scoring group than by a lower-scoring group. Its upper-minus-lower discrimination index is

☐ perfect reliability

☐ negative discrimination

☐ content validity

☐ high difficulty
5. A reliability coefficient of .95 shows that a test is valid.

☐ True ☐ False
6. Form A is administered on Monday and Form B a week later. What does this comparison change?

☐ Only the form, with time effects ruled out.

☐ Both form and occasion, so their contributions are not isolated by this comparison alone.

☐ Only the occasion, with form effects ruled out.
7. A high coefficient alpha by itself proves that every item measures one dimension.

☐ True ☐ False
8. A proposed mathematics-reasoning score correlates strongly with reading skill. Which response is best?

☐ Declare the score valid because every large correlation supports its use.

☐ Examine whether reading demands contribute irrelevant variation for this intended use.

☐ Declare all mathematical content invalid without further evidence.

9. The transformation $100 + 15z$ by itself establishes a percentile rank, regardless of the reference distribution.

☐ True ☐ False
10. An SEM of 4 points guarantees that every observed score is within 4 points of its corresponding true score.

☐ True ☐ False

1. **0.82**

$r_{\text{full}} = 2r_{\text{half}} / (1 + r_{\text{half}}) = 1.40 / 1.70 \approx 0.82$. The estimate depends on comparable halves and the measurement model; length alone does not guarantee validity.

2. **4**

$SEM = 10 \times \sqrt{1 - 0.84} = 10 \times 0.4 = 4$ points. The reliability and SD must match the intended model and scale.

3. **35**

T-scores have mean 50 and SD 10: $T = 50 + 10z = 50 + 10(-1.5) = 35$.

4. **negative discrimination**

The difference is negative. Check the key, wording, coding, sample and intended content; the sign alone does not establish one cause or require automatic removal.

5. **False**

A reliability coefficient describes consistency under a particular model. It does not establish the evidence needed for a proposed interpretation and use.

6. **Both form and occasion, so their contributions are not isolated by this comparison alone.**

The same people encounter another form on another occasion. Further design information is needed to separate those sources.

7. **False**

Alpha does not establish unidimensionality or a valid interpretation. Examine internal structure and other relevant evidence.

8. **Examine whether reading demands contribute irrelevant variation for this intended use.**

Whether the relation is expected or problematic depends on the construct and intended interpretation. Investigate content and response demands rather than using the correlation as a universal badge.

9. **False**

The transformation sets a mean/SD-style scale. Percentile rank requires the distribution or an appropriate norm table.

10. **False**

SEM describes error variation under a model, not a maximum error or a guaranteed individual interval.