Historical perspectives and psychometric interpretation

Supported GRE teaching and original transfer practice

Historical perspectives and psychometric interpretation

- Compare the explanations offered by the main perspectives in psychology.
- Separate **reliability** 信度 from **validity** 效度.
- Tell a **percentile** 百分位数 from a percentage correct, and norm-referenced from criterion-referenced scores.
- Find the **positive predictive value** 阳性预测值 of a screening test from base rates.

Perspectives

- Biological: which brain systems, genes and hormones are involved?
- Behavioural: what learning history and consequences shape it?
- Cognitive: what information processing produces it?

Perspectives (2)

- Psychodynamic: what unconscious conflicts drive it? Humanistic: how does it relate to growth and choice?
- These are complementary emphases, not rival labels for the same claim. One observation rarely settles which is "right".

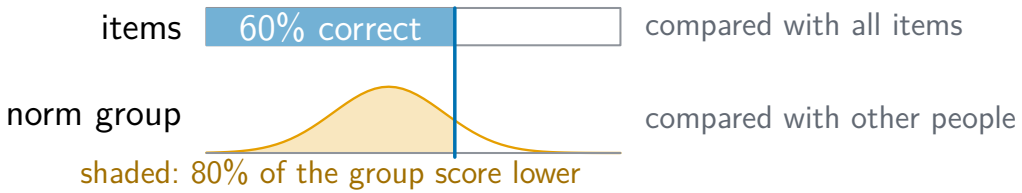
Reliability and validity

- **Reliability** is consistency: the same person gets similar scores on retesting, on another form, or from another scorer.
- **Validity** is evidence that scores support a particular interpretation and use.
- A test can be reliable without being valid: a consistent scale that measures something else. It cannot be valid for a use without being reliable.

Norms and standards

- Standardisation means the same instructions, timing and scoring for everyone, and a reference (norm) group.
- A norm-referenced score compares a person with the norm group. A criterion-referenced score compares performance with a fixed standard, such as "can solve linear equations".
- A percentile is the percentage of the norm group scoring lower. A percentage correct is the share of items answered correctly. They can differ widely.
- The norm group must fit the people tested: norms from one country or age group may mislead for another.

Figure: Norms and standards



- Percent correct compares with items; percentile compares with people.
- Match each part of the diagram to the text.

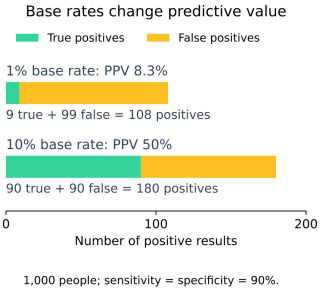
Screening and base rates

- Sensitivity: the share of people with the condition who test positive.
- Specificity: the share of people without it who test negative.
- **Positive predictive value**: the share of positive results that are correct. It depends strongly on how common the condition is.

Worked example: a rare condition

- Sensitivity is $9/10 = 90\%$, and specificity is $891/990 = 90\%$.
- There are $9 + 99 = 108$ positive results, of which only $9/108 = 8.3\%$ are correct.
- If 100 of the 1,000 had the condition, the same test would give 90 true and 90 false positives: a predictive value of 50%.

Figure: Worked example: a rare condition



1,000 people; sensitivity = specificity = 90%.

- The same test means far more when the condition is common: base rates set the predictive value
- Read the axes and units first.

Worked example: two numbers for one score

- These are not contradictory: 60% compares the score with the items, and the 80th percentile compares it with other students.
- On a hard test, many students score below 60%.
- Report which comparison is meant, and which norm group.

Common errors

- Equating sensitivity with the chance that a positive result is correct.
- Treating high **reliability** as proof of **validity**.
- Confusing a percentile with a percentage correct.

Common errors (2)

- Using norms from a group that does not match the people tested.
- Treating perspectives as competing answers to the same question.

Key terms

- **reliability**: The consistency of scores across repeated measurement.
- **validity**: Evidence that test scores support a particular interpretation and use.
- **percentile**: The percentage of a norm group that scores below a given score.
- **positive predictive value**: The share of positive test results that are correct.

Check yourself

- A scale always reads 2 kg too heavy. Is it reliable, valid, both or neither?
- A screen flags 18 of 20 true cases and 198 of 1,980 others. What fraction of positives are true cases?
- Which kind of score says whether a student can drive safely: norm-referenced or criterion-referenced?