Measurement issues

- Measurement
- Sampling
- Statistical Analysis

Types of measures

- Physiological
- Behavioral measures
- Self-reports

What measures should you use?

- Multiple measures
- Established measures

Evaluating measures (desirable properties)

- Reliability
  - Consistency
- Validity

Types of reliability

- Test-retest
- Interitem
- Interrater or interobserver
- In practice, reliability often established through replication.

Reliability is affected by…

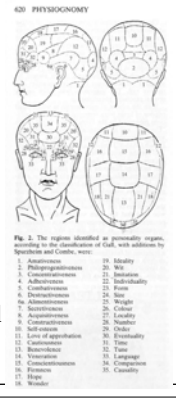
- Attention lapses
- Carelessness in recording observations
- Differences among observers on how behaviors are rated
  - Ex: the Washoe studies
- Observer biases

Reliability vs. Validity

- Reliability is about the consistency of a measure
- Validity is about whether it actually measures what it is supposed to

Invalid but reliable: Phrenology

- Popular 19th century 'science.'
- The measurement of lumps & bumps was highly reliable (since repeat measurements of the bumps would agree).
- The measurement is invalid since the bumps do not relate to any of the 35 or so supposed 'mental faculties'.



Types of validity

- Construct validity
  - Face validity
  - Criterion validity
  - Concurrent validity
  - Convergent validity
  - Predictive criterion validity
  - Discriminant validity

Construct validity

- The approximate truth of the conclusion that your operationalization accurately reflects its construct.
- An assessment of how well you translated your ideas or theories into actual programs or measures.
- Construct validity can be viewed as a "truth in labeling" kind of issue.

Assessing construct validity

- Face validity
  - extent to which a measure appears to measure what its suppose to measure.
  - the weakest way to try to demonstrate construct validity, since something can have face validity without really having validity (as in phrenology).
- Criterion validity
  - check the performance of your operationalization against some criterion.

Criteria for validity

- Concurrent validity
  - Assess the operationalization's ability to distinguish between groups that it should theoretically be able to distinguish between.
  - Example: New test for SLI should distinguish between kids who have SLI and kids who have ADD
- Convergent validity
  - Examine the degree to which the operationalization is similar to (converges on) other operationalizations.
  - Our test should correlate with other tests of same topic.

Criteria for validity, cont.

- Predictive criterion validity
  - Assess the operationalization's ability to predict something it should theoretically be able to predict.
  - Example: a measure of math ability should be able to predict how well a person will do in an engineering-based profession.
- Discriminant validity
  - Examine the degree to which the operationalization is not similar to (diverges from) other operationalizations that it theoretically should be not be similar to.
  - Example: our test of arithmetic skills should not correlate with scores on tests for verbal ability.

Other measurement problems

- Reactivity
- Demand characteristics
- Observer Bias
- Ceiling & floor effects

Reactivity

- Influence that an observer has on the behavior under observation.
- Includes
  - social desirability
  - pleasing the experimenter

Demand characteristics

- Cues used by subjects to guide their behavior.
- Example 1: if the experimenter frowns after certain answers, or nods after different ones, that may tell the subject that there are certain "good" responses vs. "bad" responses.

Clever Hans

- A horse that lived in Germany in the early 1900s.
- His owner, Wilhelm von Osten, claimed that his horse could answer a wide variety of questions, such as solving mathematical problems and telling the time, and communicate the answers using hoof-taps.
- The horse performed "almost as well" when von Osten was absent as when the master was present - the effect was not trickery.

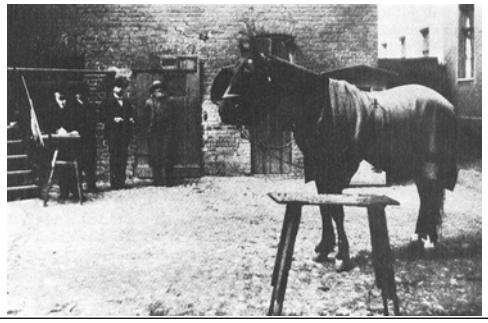
Image: www.kbrhorse.net/tra/hans.html

Clever Hans, cont.

Clever Hans, cont.

- Scientist Oskar Pfungst uncovered Hans's one weakness: he was unable to respond correctly when no one in front of him knew the answer to the question at hand.

Clever Hans, cont.



Clever Hans, cont.

- No one was tipping off Clever Hans intentionally.
- The horse had learned to identify subtle tensing and relaxing of muscles that occur in someone who is anticipating the correct answer.
- Hans would tap his hoof until he saw the subconscious twitch in observers who knew he had arrived at the right spot in the alphabet, and there Hans would stop.

Observer Bias

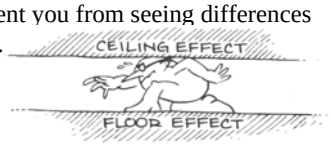
- Example: Study of mental institutions
  - Individuals misrepresented their names & occupations and sought admission to different mental institutions, claiming they heard voices.
  - After being admitted, they stopped complaining of symptoms and just acted normal.
  - But once these people were originally labeled as schizophrenic, their behavior was interpreted in light of this label.
  - This bias prevented the staff from detecting the pseudopatient's sanity.

Observer bias

- Expectancy effects
- To avoid this, the observer should be <u>blind</u> to condition.

Ceiling and floor effects

- Ceiling effect:
- Floor effect:
- These can prevent you from seeing differences that really exist.



Martin, D.W. (1985) Doing Psychology Experiments, 2nd Ed., p. 83

Example

- If you make a really easy language task, you may not see differences between kids with SLI and kids without -- all will be perfect.
- If you make a really tough perception in noise task, where everyone does very poorly, you may not see differences between individuals with and without hearing impairments.
- These do not mean the groups do not really differ.

Ceiling effects

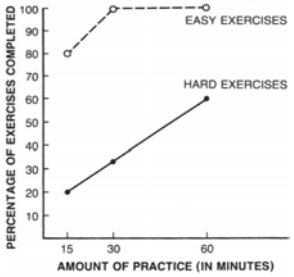
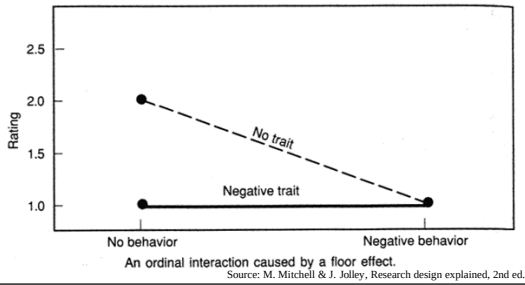


Illustration of a ceiling effect.

Source: J.J. Shaughnessy & E.B. Zechmeister, Research Methods in Psychology, 2nd ed.

- This may look like an interaction, and statistically an interaction might come out.
- But the interaction may simply be the result of the ceiling effect.

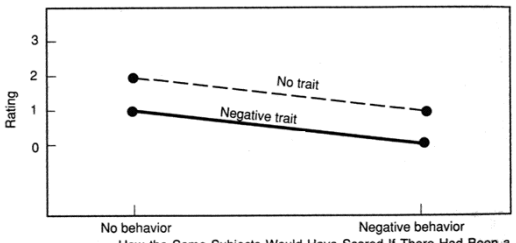
Interaction caused by floor effect



An ordinal interaction caused by a floor effect.

Source: M. Mitchell & J. Jolley, Research design explained, 2nd ed.

No interaction when floor is lowered



How the Same Subjects Would Have Scored If There Had Been a "Lower Floor"

**Note that in the second graph, the lines are parallel and the "interaction" disappears.*

Source: M. Mitchell & J. Jolley, Research design explained, 2nd ed.