

Psy 427
Cal State Northridge
Andrew Ainsworth PhD

TEST ADMINISTRATION

Psy 427 - Cal State Northridge

FACTORS AFFECTING TEST ADMINISTRATION

- × When we talk about reliability, we are interested in random sources of error.
+ Observed Score = True Score + Error
- × When tests are actually administered, however, there are other sources of error aside from random error.

2

Psy 427 - Cal State Northridge

FACTORS AFFECTING TEST ADMINISTRATION

- × Rapport
- × Ethnicity
- × Language
- × Training of Test Administrators
- × Expectancy Effects
- × Use of Reinforcers
- × Computer-Assisted Testing
- × Subject Variables

3

RAPPORT

- × Importance of establishing rapport
 - + Feldman & Sullivan (1960) - WISC
- × Enhanced rapport vs. neutral rapport younger children (through grade 3) did not benefit from enhanced rapport
- × Older children (grades 5-9) produced higher IQ scores under enhanced rapport
 - + Enhanced Rapport = mean IQ of 122
 - + Neutral Rapport = mean IQ of 109

4

RAPPORT

- × Children score lower on IQ test when the administrator made disapproving comments ("I thought you could do better) then when administrators made neutral or positive comments (Witmer, Bernstein and Dunham, 1971)
- × Children unfamiliar with the administrator did significantly worse on a reading test compared to children familiar with the administrator (DeRosa and Patalano, 1991)

5

RAPPORT

- × Importance of establishing rapport
 - + Fuchs & Fuchs (1986) - meta-analysis
 - × 22 studies involving 1489 children
 - × 4 IQ point increase when the examiner was familiar with the test taker, in general
 - × 7.6 IQ point increase when familiarity and lower SES co-occurred

6

RAPPORT

- × Importance of establishing rapport?
 - + Self-report vs. interview of attitudinal surveys
 - × People disclose MORE on self-report than they do to interviewers
 - × People disclose MORE to computers than they do to human interviewers
- × Conclusions:
 - + Rapport is important in situations that are not viewed as "personal" or those typically subject to social desirability.

ETHNICITY

- × Should children of one ethnicity be tested only by test administrators of the same ethnicity?
- × Majority of studies have found nonsignificant effects for cross-ethnic administration of most intelligence tests.
- × The only significant findings have been when paraprofessionals have administered the tests.
- × Why no differences?
 - + Standardized procedures

LANGUAGE

- × How valid are tests given in English to bilingual or Limited-English Proficient (LEP) individuals?
- × What about translating tests?
- × Language
 - + Standard of practice: administer a test in the most proficient language.
 - + BUT - what about the normative sample?
 - + How comparable are the scores from these individuals?
 - + Can IRT help?
- × Interpreters: another potential source of bias

TRAINING

- ✘ Administration and scoring errors are a large source of bias.
 - + Typical graduate training: 2-4 administrations of a test (in class)
 - ✘ importance of fieldwork placements
 - ✘ majority of testing practice obtained in fieldwork placements
- ✘ Error rates on WAIS administrations decrease after 10 administrations(!)

EXPECTANCY EFFECTS

- ✘ Also known as: Rosenthal effects
 - + Robert Rosenthal, Harvard University
 - + Subjects perform in a manner consistent with experimenter's (test administrator's) expectations
 - ✘ works with humans, works with rats
 - + Effects not limited to experiments, also occurs on standardized tests
 - ✘ students asked to score ambiguous responses will give more points to people they like, or think are bright.
 - ✘ People find what they expect

EXPECTANCY EFFECTS

- ✘ Expectancy and test administration
 - + Rosenthal - expectancy effects are triggered by non-verbal cues, and the experimenter/ administrator may not even be aware
 - + Expectancy effects have small and varied influence on test outcomes; careful study is required

USE OF REINFORCEMENT

- ✘ No clear and consistent difference between studies using reinforcement showing positive or negative effects.
- ✘ Individual studies, however, are compelling
 - + Terrell, Taylor, & Terrell (1978) found a 17.6 point increase in IQ scores when African-American children were given culturally appropriate feedback by African-American test administrators.

USE OF REINFORCEMENT

- ✘ General guidelines:
 - + Check with the testing manual first
 - + Generally OK to reward EFFORT, not answers.

COMPUTER-ASSISTED TEST ADMINISTRATION

- ✘ Advantages
 - + The obvious connection to Item Response Theory and the ability to tailor tests to a persons ability
 - + Highly Standardized
 - + Precision of Timing
 - + Lessened Dependence on Human Testers
 - + Pacing (no need to rush respondents)
 - + Control of Bias (from the test administrator, etc.)

COMPUTER-ASSISTED TEST ADMINISTRATION

- ✘ Computer adaptive versions of tests have shown no large differences between computer assisted and paper-and-pencil versions
- ✘ Computer versions can be more accurate and take less time (e.g. IRT and CAT)
- ✘ Some people enjoy the computer format and even prefer it

COMPUTER-ASSISTED TEST ADMINISTRATION

- ✘ One study (Locke and Gilbert, 1995) showed that when respondents are asked about sensitive material (e.g. MMPI, drinking, etc.) they were more honest when the tests were administered via computer vs. questionnaire or interview.
- ✘ CAT has been applied to the MMPI, personnel selection and cognitive tests among others

COMPUTER-ASSISTED TEST ADMINISTRATION

- ✘ The big concern with computer aided testing is that it will lead to the computer generated reports landing in the wrong (inexperienced) hands and misinterpreted

SUBJECT VARIABLES

- ✘ The state of the subject can also be a source of error when administering a test
 - + Illness
 - + Insomnia
 - + Test-anxiety
 - + Drugs (prescription and recreational)
 - + Hormones (e.g. menstruation) – variations in perceptual motor coordination varied with cycle (better away from menses; effects reverse for other tasks)

FACTORS AFFECTING BEHAVIORAL ASSESSMENT

- ✘ Issues that arise when people (judges) act as the testing instrument
- ✘ Human judges are subject to problems that add to the error when assessing respondents
 - + Reactivity
 - + Drift
 - + Expectancies (same as with test administration)
 - + Deception

REACTIVITY

- ✘ The reliability of behavioral assessments is usually assessed using inter-rater reliability (consistency among raters) or by having supervisor make periodic checks
- ✘ Reliability tends to be highest when the judges know they are being evaluated either by supervisor or against one another
- ✘ This increase is called REACTIVITY

REACTIVITY

- ✘ One study (Reid, 1970) showed that the reliability of observers rating decreased 25% when they were told that they would not be compared to a standard
- ✘ Many studies report inter-rater reliability, but care should be exercised because these numbers typically are calculated during training and drop during the administration

DRIFT

- ✘ Evaluators (i.e. judges, observers) are typically trained and given a strict set of rules in which to follow when evaluating targets
- ✘ After some time the evaluators may not be following as strict a set of guidelines as they were trained to follow; this is called DRIFT
- ✘ To avoid this, evaluators should be periodically retrained on the assessment criteria

DECEPTION

- ✘ How do you assess someone who does not want to be "accurately" assessed?
- ✘ People in general are awefull at detecting lying
 - + Secret service agents only scored above chance (Ekman and O'Sullivan, 1991)
- ✘ Lie detection as an industry
 - + Lie detectors (even though questionable at best)
 - + Commercial tests of honesty and integrity (questionable validity)
