

Observation & Measurement

Claudia J. Stanny
PSY 6217 – Research Design



Choosing Variables to Measure

- ☞ **Research tradition**
 - ♦ *Important of the literature review*
 - ♦ *Focus on current preferences in measures*
- ☞ **Theory**
 - ♦ *Direct attention to particular measures*
- ☞ **New technology**
 - ♦ *Physiological measures enabled by changes in technology (EEG, ERPs, PET, MRI, fMRI)*
 - ♦ *Some times these create new research questions*
- ☞ **Availability**
 - ♦ *Do you have access to equipment needed for the measure?*
 - ♦ *Do you have the skill and/or training needed for using this measure?*

Reliability of a Measure

- ☞ **Reliability as consistency**
- ☞ **Physical measures**
 - ♦ *Precision (size of units)*
 - ♦ *Range of variation possible with repeated measures*
- ☞ **Population estimates**
 - ♦ *Margin of error (e.g., in reporting public opinion)*

Reliability of a Measure

- ☞ **Judgments or ratings by observers**
 - ♦ *Interrater reliability*
 - ♦ *Aggregate reliability of multiple judges*
 - ♦ *Percent agreement (distortions with high frequency categories)*
 - ♦ *Cronbach's alpha*
- ☞ **Psychological tests or measurements**
 - ♦ *Internal consistency*
 - ♦ *Kuder-Richardson (K-R 20)*
 - ♦ *Cronbach's alpha applied to test items as judges*
 - ♦ *Spearman Brown*
 - ♦ *Test-retest reliability*
 - ♦ *Parallel form reliability*
 - ♦ *Split-half reliability*

Accuracy of a Measure

- ☞ **Agreement with a known standard**
 - ♦ *Does such a standard exist?*
 - ♦ *Most likely for physical measures*
 - ♦ *Problems with establishing a "gold standard" measure for psychological variables*
- ☞ **Interaction of accuracy and precision**

Validity of a Measure

- ☞ **Face validity**
- ☞ **Content validity**
- ☞ **Criterion-based validity**
 - ♦ *Concurrent validity*
 - ♦ *Predictive Validity*
- ☞ **Discriminant validity**
- ☞ **Construct validity**
 - ♦ *Combination of criterion-based and discriminant validity*
 - ♦ *Validity of the underlying construct as well as the measure is at stake*

Scales of Measurement

- ☞ **Nominal Scales (Categorical Data)**
 - *Identify differences*
- ☞ **Ordinal Scales (Rank Order Data)**
 - *Identify differences*
 - *Order by magnitude*
- ☞ **Interval Scales**
 - *Identify differences*
 - *Order by magnitude*
 - *One unit of change has the same meaning across the scale*
 - *Zero is an arbitrary value on the scale*
- ☞ **Ratio Scales**
 - *Identify differences*
 - *Order by magnitude*
 - *One unit of change has the same meaning across the scale*
 - *“True Zero” -- Zero means complete absence of the trait measured*

Relevance of Measurement Scales

- ☞ **Constraints on statistical analysis**
 - *Nominal – Non-parametric statistics*
 - *Ordinal – Special non-parametric statistics for rank order data*
 - *Interval & Ratio – Parametric statistics*
- ☞ **Do the statistics “care” about how the numbers were generated?**
- ☞ **How should numerical findings be interpreted?**
- ☞ **Ecological validity of the measure**
 - *A confidence scale gives more information than a yes/no decision*
 - *Juries make guilty/not guilty verdicts*
 - *Managers make hire/do not hire decisions*

Adequacy of a Measure

- ☞ **Sensitivity to change in the underlying variable**
 - *are you measuring behaviors that respond to your manipulations?*
 - *e.g., choosing between accuracy and reaction time*
- ☞ **Does the measure produce enough variability in responses to detect effects of manipulations?**
 - *Restriction of range problems*
 - *Ceiling effects*
 - *Floor (basement) effects*

Varieties of Measures

- ☞ **Behavioral measures**
 - *Frequency, latency, duration, number of errors*
- ☞ **Physiological measures**
 - *Non-invasive measures*
 - *EEG, GSR, heart rate, blood pressure*
 - *Invasive measures*
 - *Single-unit recording, blood analysis*
 - *Calibration issues*
- ☞ **Self-report measures**
 - *Q-sorts, Likert-type scales, ratings, diary entries, questionnaires & surveys*
 - *Problems of reliability, validity, & reactivity*

Criteria for a Good Measure

- ☞ **Reliable**
- ☞ **Valid**
- ☞ **Accepted as an established measure**
- ☞ **Quality of information provided**
- ☞ **Sensitive to important changes in the underlying construct to be measured**
- ☞ **Acceptable range of values for use in statistical analysis**
- ☞ **Suitable for use with the subject population**

Problems of Reactivity

- ☞ **Demand Characteristics**
- ☞ **Role Attitudes of participants**
 - *Cooperative*
 - *Apprehensive or Defensive*
 - *Negative*
- ☞ **Reactivity of local events**
- ☞ **Experimenter Effects**
 - *Characteristics of the experimenter (biosocial)*
 - *Expectations of the experimenter*

Control of Reactivity

☞ **Demand Characteristics**

- *disguise the research hypothesis*
- *direct attention away from critical measures*

☞ **Expectancy Effects**

- *Single-blind controls*
 - *Participants are blind (demand characteristics)*
 - *Experimenters or scorers are blind*
- *Double-blind controls*
 - *Participants & researchers are blind*
- *Automation of procedures*

Automation of Experimental Procedures

☞ **Advantages**

- *Precision in control of timed events*
- *Reliability of data recording*
- *Consistency of treatment of participants*
- *Potential reduction of expectancy effects*
- *Potential to collect data from many people at the same time*

☞ **Disadvantages**

- *Do participants understand the instructions?*
- *Are participants motivated to do the task properly?*
- *Are participants consulting one another?*

Conducting Research on the Internet

☞ **Multiple submissions**

☞

Detecting and Correcting Problems

☞ **Pilot Studies**

☞ **Manipulation Checks**

☞ **Debriefing**