

# Lecture 19: Systematic Observation

Dewayne E Perry

ENS 623

perry@ece.utexas.edu

# Categorizing Judgments

- Choice of a given method of data collection depends on the context to be used
- Eg, making participant observer records
  - ↪ Rosenhan 73 - stigmatization of mentally ill
  - ↪ Volunteers admitted themselves to mental hospitals
  - ↪ Kept detailed records of interactions with psychiatrists, psychologists, etc
  - ↪ Involves making judgments about actions and conversations
  - ↪ Conclusion: staff seem to avoid interacting with mental patients
  - ↪ Actors (ie, psychs etc) were not told they were being observed
    - More typically true that actors know of observations
    - And may be more selective about cooperation for privacy reasons
    - Observer must be attuned to this limitation

# Categorizing Judgments

## → Unobtrusive methods

- ↪ Use of physical traces

- ↪ Use of archival methods

  - Content analysis: method of categorizing judgments

  - Will look at general guidelines for CA and judge selection

- ↪ Non-reactive observations

## → Rating scales

- ↪ Among most widely used method for categorizing judgments

- ↪ Traditional rating scales

- ↪ Classifier vs evaluator

# Observational Records

- Do not yet have a tradition of how to formulate the most objective results
- Problem has surfaced about objectivity: eg, Margaret Mead and others
- Seem to be more attuned to the persistent and unique limitations

# Observational Records

## → Systematic observation

- ↳ **Conforms to a definite plan of action**
  - What is to be observed, ignored, recorded
  - What measurements and instruments to be used
- ↳ **Frequently work in teams and employ checks and balances**
  - To control for biases
- ↳ **Try to establish rapport by framing questions in local context**
  - Cf, fault study, time studies
- ↳ **Deal with observer as artifact using concealment**
  - Observer bias or interference

# Observational Records

→ Produce meticulous records: ethnographic records

↪ Field notes, pictures, tapes, etc

↪ Unit of analysis is *activity*

➤ an intentional act, behavior or incident

➤ Aimed at affecting the status of events

↪ What is happening, what are people doing

↪ 6 questions to consider

➤ What is the purpose of the activity

➤ What procedures are used

➤ What are the time and space requirements

➤ What are the personal requirements

➤ What are the occasions for performance

# Accurate Field Notes

- Important to indicate verbatim quotes vs normal language interpretation
- People tend to translate and simplify what they see and hear
- Different kinds of accounts
  - ↳ Condensed accounts
    - Phrases, single words, unconnected sentences
    - Embellished by observers own transcriptions
  - ↳ Expanded account
    - Researchers additions to the condensed account
  - ↳ Fieldwork journal
    - Diary record of personal or subjective side of field work
  - ↳ Grand tour vs mini tour observations
    - General impressions based on some or all of the following
      - ✓ Space, actors, activity, objects, act, event, time, goals, feelings, etc
    - Concrete details of a unit of experience - in depth

# Choosing Judges

- Regard each judge or rater as interchangeable
  - ↪ Do not consider differences between them
  - ↪ Interested in them being equivalent
- Possible ways to select
  - ↪ Common characteristics -
    - eg, college grads, domain experts
  - ↪ On the basis of pilot testing
    - Identified specific characteristics for intended judgments
  - ↪ According to accuracy on some relevant criterion
    - Begin with a pool of judges and test
- How do we select really accurate judges
  - ↪ Good basis of literature
    - On correcting response measures for guessing and partial information
  - ↪ Estimating guessing on tests:
    - $RI \text{ adjusted} = RI - [WI/(A-1)]$ 
      - ✓ RI - right items
      - ✓ WI - wrong items
      - ✓ A - number of alternatives

# Checklists

## → 3 steps to accurate information

- ↳ Define each category of behavior that is relevant as specifically as possible
- ↳ Decide where and when the observations will be made
- ↳ Select and train judges in the use of the scales to record the particular variables under investigation

# Checklists

## → Bales 50; Bales & Cohen 79

### ↪ Problem solving

- Eg, consider using this approach to look at informal interactions among SW developers

### ↪ 15-20 codable interactions per minute

### ↪ Half involve problem solving; half positive or negative questions

### ↪ First third:

- Giving information tends to be the most frequent behavior

### ↪ Last third:

- Offering suggestions

## → Kipnis 69 70 76

### ↪ Tactics of power

- Strong tactics: Other viewpoint as foolish
- Rational tactics: Stating view and let others decide
- Weak tactics: Flattery, modeling, pretending

# Unobtrusive Indicators

- Behavior so complicated no agreed upon way of studying or measuring it
- Disguised measures, archives, physical traces
- Webb et al - seminal work
  - ↳ Objection that investigators are overly dependent on a single criterion
  - ↳ Underscore systematic pluralism - triangulate
- Unobtrusive measure problem:
  - ↳ Arguably less precise than direct
  - ↳ Advantage: non-reactive instruments
- Wrightsman 69 - tax stickers and bumper stickers

# Physical Traces

→ Webb and coauthors give a virtual compendium of unobtrusive measures

↳ Physical clues

➤ eg, radio buttons for advertising targets

↳ Wear and tear on library books

↳ Nose prints on museum exhibits

↳ Language behavior

➤ Content of messages

→ Archives

↳ Actuarial records

↳ Political and judicial records

↳ Other governmental records

↳ Mass media

↳ Sales records

↳ Industrial and institutional records

↳ Various other written documents

# Physical Traces

## → Important to be sensitive to data transformations

↪ Eg, actual numbers relative to size increase over time

↪ Eg, percentages instead of frequencies

## → Documentary research

↪ Content analysis of historical documents

↪ Establish authenticity of documents

↪ Problems: availability, sampling, truth or not

↪ Platt 81 for methodological guidelines

## → Simple observation

↪ Length of hair and methodological disposition (tough/soft mindedness)

↪ Tough minded had shorter hair than soft minded

## → Contrived observation

↪ Introduces variables of interest

↪ Watches what happens

↪ Eg, metaphorically cat among the pigeons

# Content Analysis

- Checklists and tally lists are used in content analysis
- Objective and systematic strategy of
  - ↳ decomposing messages
  - ↳ evaluating and classifying their content
  - ↳ to reveal their specific characteristics
- Berelson 52 - definitive work
  - ↳ Content analysis in communication research
  - ↳ make inferences from textual material from counting words, sentences, ideas, categories etc

# Content Analysis

## → Guidelines

- ↳ **Analyses should be consistent among judges**
  - Different coders should produce the same results
  - Differences affect reliability
  - Careful definition + proper training -> high reliability
- ↳ **Essential that categories be relevant to hypotheses**
  - What is the communication about, what is said
  - Consider alternative units before selecting
- ↳ **Important to decide on good sampling procedure**
  - CA is time consuming
  - Important that material is representative to justify effort

# Content Analysis

## → Advantages

- ↪ Requires little more than common sense to develop coding system and then to implement it
- ↪ It a shoestring methodology - labor intensive but capital minimum
- ↪ Safe methodology - can add information over time if missed or incorrectly coded
- ↪ Forces scrutinization of material being evaluated and classified

# Rating scales

→ Rating scales impart a sense of structure on the response

- ↳ Numerical scale
- ↳ Forced choice scale
- ↳ Graphic rating scale

→ Numeric Scale

- ↳ A sequence of defined numbers
  - May be explicit or implicit (via some ordering)
  - Worry about ambiguity and complexity
  - Worry whether to include middle category or force selection
  - Easiest to construct and use
  - Simplest in terms of data analysis
  - More vulnerable to many biases and errors than other forms
    - ✓ Eg, halo effect

# Rating Scales

## → Forced-Choice Scale

- ↪ Eg, X is energetic or X is intelligent
  - Forces choice of one
- ↪ Ranges from two statements to N
- ↪ Range from select one to select most and least descriptive
- ↪ Time consuming to develop
- ↪ Sometimes resisted to by subjects being forced to choose
  - Between equally favorable or unfavorable

## → Highland and Berkshire 51

- ↪ Extensive analysis of 6 forms of instruments
- ↪ A: 2 statements per item
  - ++ or -- select more/less descriptive
  - One of most popular
- ↪ B: 3 statements per item
  - +++ or --- select most/least descriptive
  - One of least tendency to bias
  - One of least popular

# Rating Scales

- ↪ **C: 4 statements, all favorable**
  - +++++ select two most descriptive
  - Best choice in many situations
  - Tended to give highest validity results (with D)
  - Least tendency to bias (with B)
- ↪ **D: 4 statements, all favorable**
  - +++++ select most and least descriptive
  - Tended to give highest validity results (with C)
  - One of least popular
- ↪ **E: 4 stmts, 2 favorable, 2 unfavorable**
  - ++-- select most and least descriptive
  - Yielded highest coefficient of reliability (with F)
- ↪ **F: 5 stmts, 2 fav, 2 unfav, 1 neutral**
  - ++0-- select 1 most and least descriptive
  - Yielded highest coefficient of reliability (with E)

# Rating Scales

## → Multiple Points

### ↳ Graphic scale

- Usually a straight line, either H or V
- Positive, neutral, negative
- Problem with previous two: units are equidistant

### ↳ How many points needed?

- Study showed internal consistency reliability increased steadily up to a 7 point scale but not beyond
- Anchor points need to be simple, unidimensional and unambiguous

# Rating Errors and Control

- Assumption: rater is capable to an acceptable degree of of precision and objectivity
- One type of bias: halo effect
  - ↳ Judged favorably on one trait tends towards favorable on others also
  - ↳ Knowledge of previous outstanding performance or trait -> effects current rating
  - ↳ Most prevalent when characteristic
    - Is not easily observable, or not clearly defined
    - Involves relations with other people
    - Is of moral importance
- Error of leniency
  - ↳ Familiarity or ego-involved rates more positively
  - ↳ May lean over backwards and make opposite error
  - ↳ Might counteract with appropriate rating scale

# Rating Errors and Control

## → Error of central tendency

- ↪ Hesitation to give extreme scores
- ↪ Might control via expanded scale
- ↪ Eg, use a 7 point scale instead of 5

## → Logical error in rating

- ↪ Observers logic relatedness differs
- ↪ Perceptions cause problems
- ↪ Must construct as precise definitions as possible

## → Most effective strategy:

- ↪ Multiple judges who have been carefully trained
- ↪ Pool their ratings over many different trials
- ↪ Using similar and different rating methods

## → Training, practice and discussion