Psych 818
Measurement Types & Construction

Readings...
- Dawis (1987)
  - Test Construction
- Reckase (1996)
  - Test construction
- Roberts et al (1990)
  - Scaling
- Friedman & Amoo (1999)
  - Rating the rating scales

Do you know your level?
- Levels of analysis and measurement
  - A fundamental decision of any measurement is to select a level of analysis
  - This decision is often made implicitly based on existing research norms
  - Consideration of levels of analysis can greatly improve your conceptualization and measurement

Levels of Analysis

- Social
- Behavioral
- Cognitive
- Neural
- Biological
- Chemical & Physical

Levels Principles
- Mesarovic, Macko, & Takahara (1970)
  1. Level selection depends upon the investigator's knowledge and interest in the operation of the system
  2. Contexts in which the system operates on different levels are not, in general, mutually related
  3. The principles or laws used to characterize the system at one level cannot be derived from the principles used at other levels

Levels Principles
- Mesarovic, Macko, & Takahara (1970)
  1. Asymmetrical interdependence between the functioning of a system at different levels
    - Higher levels constrain the functioning of lower levels
  2. Each level has its own set of terms, concepts, and principles
  3. Understanding is increased by crossing strata
    - Lower = how
    - Higher = why
Levels Principles

- Systems differ in decomposability
  - Near decomposability - Simon
    the degree to which the behavior of a system at any one level is free of the interactions on a lower level and the degree which its interactions are irrelevant to the higher levels of the system
- "Scientific knowledge is organized in levels, not because reduction in principle is impossible, but because nature is organized in levels, and the pattern at each level is most clearly discerned by abstracting from the details of the levels far below"

Psychological Measurement Crosses the levels

- Psychology is a very broad field of investigation
  - Hormones, chemicals, genes to ...
  - Politics of nations

Physiological

- Hormones (estrogen), Neurohormones (dopamine) & Neurotransmitters (Serotonin)
- Genes
- Circadian cycles
- PET Scans
- CAT scans
- EEG (electroencephalograph)
- fMRI
- Pupilometer

Interviews and Judgment

- Job Interviews
  - Structured vs. Unstructured
- Clinical Interviews
- Observer Attributions
- Jury deliberation
- Often categorical
- Dawes – Mechanical vs. judgmental accuracy
Ability & Achievement
- Multiple Choice
- Constructed Response
- Physical Abilities
  - A tale of Firefighters and Ice skaters...

Self-Report
- Personality
- Experiences & Biodata
- Attitudes
- Beliefs
- Affect?
  - Cognitive process?
    - Nisbett & Wilson (1977). Telling more than we can know.

Other Report
- Peers
- Significant others
- Family
- Teachers
- Supervisors

Opportunity to observe slices of the person
Attribution problems
The “other” also has a personality and fallible memory

Observational
- Narratives
  - attempts to record as much as possible of what happens within
  - the focus of the observation.
  - open-ended, flexible & rich
  - time consuming to both record and interpret
  - Diary description
    - a chronological record of behavior made after the behavior occurs
      - Anecdotal record
        - detailed notes on an identified situation, recorded while the behavior is occurring
      - Specimen description
        - a descriptive narrative, recorded after the behavior occurs
  - Running record
    - a sequential record over a given time recorded while the behavior is occurring

Observational
- Time Sampling
  - an observation of what happens within a given period
    - of time, coded with tallies or symbols while the behavior is occurring, used to document
    - frequency of specific behaviors.
    - more objective than narrative records, less time consuming, and offers a way to observe and record behavior
      - of more people simultaneously.
    - closed-ended, limited to what happens in the specified time
      - interval, and lacking in behavioral and contextual detail.

Observational
- Event Sampling
  - an observation of an event that has been
    - defined in advance and what happens before
    - and after, recorded briefly while it is taking place
    - objective
      - event sampling is closed-ended and limited

Other Report
- Opportunity to observe slices of the person
- Attribution problems
- The “other” also has a personality and fallible memory
Experience Sampling

- Advantages
  - Quantitatively study everyday experiences, moods, and behaviors in their natural context
  - Measures administered at random or scheduled intervals during the activity of interest to researchers
  - Less subject to bias in recall of events, mood, behaviors, or other psychologically important variables

Experience Sampling

- Disadvantages
  - Intrusiveness
  - Missing data & lack of commitment
  - Interruptions
  - Hawthorne effects?
  - Huge amounts of data

Internet Assessment

- Equivalence?
  - Time of day
  - Context
  - Security
  - Breaks or pauses
  - Technological problems - missing data

Measurement Mistakes

- Phrenology

More Mistakes

- Craniochemistry...
  - Gould’s Phrenology of man
  - Race - demonstrated that women are inferior to men
  - Larger brain means greater potential
effective intelligence
  - Favoring men for political positions
  - Effectiveness of societal values on science
  - Measures are like handguns - lacking awareness and responsibility

More Mistakes

- Polygraph
### General principles

1. vocabulary level – should match the population
2. use reading level algorithms to check
3. avoid slang – differences in familiarity
4. avoid absolute terms – difficult to defend due to exceptions
5. avoid conjunctions
6. avoid absolute terms – difficult to defend due to exceptions
7. use non-sexist language – pronouns and possessive verbs for both sexes
8. brevity – item stem and alternatives brief with all necessary information
9. one content topic per item – no compound items
10. Items should cover domain of construct

### Probable Mistakes

- Penile Plethysmography
  - Started in Czechoslovakia to detect sexual preference
  - Now used in pedophilia cases
  - Major issues with reliability and fakability (validity)

### Constructing Alternatives

1. 4 or 5 alternatives per item – one correct and 3 or 4 distractors
2. all alternatives should be plausible, especially the distractors
3. all alternatives should be plausible, especially the distractors
4. one best answer
5. care of evidence for incorrect alternatives
6. obtain opinion of experts
7. document source of correct answer

### Constructing Multiple Choice and Self-Report Measures

1. Clear Construct Definition
   - What is it and what isn't it!
   - Identify and define sub-dimensions
   - goal orientation example
2. Specification Chart
   - Relative importance of sub-dimensions
   - Number & type of items per dimension
3. Identify Population of Interest
4. Review existing measures/interview sample
5. Select scaling method
6. Write Items
7. Pilot Test & Item Analysis
8. Item Revision

### Constructing Alternatives

1. 4 or 5 alternatives per item – one correct and 3 or 4 distractors
2. all alternatives should be plausible, especially the distractors
3. all alternatives should be plausible, especially the distractors
4. one best answer
5. care of evidence for incorrect alternatives
6. obtain opinion of experts
7. document source of correct answer

### Writing the Item Stem

1. Don't include implicit assumptions in the stem (e.g., how concerned are you about…)
2. one clear, central problem completely
3. all alternatives should be plausible, especially the distractors
4. brevity – item stem and alternatives brief with all necessary information
5. one content topic per item – no compound items
6. Items should cover domain of construct
Constructing Alternatives

**Direct Clue Errors**

- **more obvious, require more common sense**

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical sequence</td>
<td>A list of numbers, possibly with a pattern.</td>
</tr>
<tr>
<td>Overlapping</td>
<td>Two statements that are very similar.</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>A statement that is not clear or unambiguous.</td>
</tr>
<tr>
<td>Oversimplification</td>
<td>A statement that is too simple or overly simplified.</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>A statement that is not clear or unambiguous.</td>
</tr>
<tr>
<td>Oversimplification</td>
<td>A statement that is too simple or overly simplified.</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>A statement that is not clear or unambiguous.</td>
</tr>
<tr>
<td>Oversimplification</td>
<td>A statement that is too simple or overly simplified.</td>
</tr>
</tbody>
</table>

**Indirect Clue Errors**

- **less obvious, require more common sense**

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept split</td>
<td>The concept is not evenly distributed among the alternatives.</td>
</tr>
<tr>
<td>Overlapping</td>
<td>Two statements that are very similar.</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>A statement that is not clear or unambiguous.</td>
</tr>
<tr>
<td>Oversimplification</td>
<td>A statement that is too simple or overly simplified.</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>A statement that is not clear or unambiguous.</td>
</tr>
<tr>
<td>Oversimplification</td>
<td>A statement that is too simple or overly simplified.</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>A statement that is not clear or unambiguous.</td>
</tr>
<tr>
<td>Oversimplification</td>
<td>A statement that is too simple or overly simplified.</td>
</tr>
</tbody>
</table>

**Constructing Alternatives**

- Do not use "all" and "none" alternatives in the same test.

**Constructing Alternatives**

- More direct clues
  - Some need for direct clue or sequence
  - One alternative only

**Constructing Alternatives**

- Correct answer in same position and equal number of times
- Position of correct response should be randomized
- Numerical sequences should be in non-random order

**Constructing Alternatives**

- Direct clues
  - More obvious, require more common sense

**Constructing Alternatives**

- Indirect clues
  - Less obvious, require more common sense

**Constructing Alternatives**

- Constructing Alternatives
  - More direct clues
    - Some need for direct clue or sequence
    - One alternative only

**Constructing Alternatives**

- Item Response Formats
  - Know what your anchors mean to respondents
  - Don't force a response (e.g., 4 point scales)
  - Respondents will express an attitude or belief that they don't actually hold
  - Scale should be balanced (equal number of good and bad scale points)
  - Understand that the asymmetry of comparisons (do you prefer x to y is not the same as y to x)
Psych 818
Measurement Types & Construction
Readings...

- Dawis (1987)
  - Test Construction
- Reckase (1996)
  - Test construction
- Roberts et al (1990)
  - Scaling
- Friedman & Amoo (1999)
  - Rating the rating scales
Do you know your level?

- Levels of analysis and measurement
  - A fundamental decision of any measurement is to select a level of analysis
  - This decision is often made implicitly based on existing research norms
  - Consideration of levels of analysis can greatly improve your conceptualization and measurement
Levels of Analysis

Adapted from: Newell (1990) - Unified Theories of Cognition
Levels Principles

Mesarovic, Macko, & Takahara (1970)

1. Level selection depends upon the investigator’s knowledge and interest in the operation of the system

2. Contexts in which the system operates on different levels are not, in general, mutually related

3. The principles or laws used to characterize the system at one level cannot be derived from the principles used at other levels
Levels Principles

Mesarovic, Macko, & Takahara (1970)

1. Asymmetrical interdependence between the functioning of a system at different levels
   - Higher levels constrain the functioning of lower levels

2. Each level has its own set of terms, concepts, and principles

3. Understanding is increased by crossing strata
   - Lower = how
   - Higher = why
 Levels Principles

- **Systems differ in decomposability**
  - **Near decomposibility - Simon**
    - the degree to which the behavior of a system at any one level is free of the interactions on a lower level and the degree which its interactions are irrelevant to the higher levels of the system
    - “Scientific knowledge is organized in levels, not because reduction in principle is impossible, but because nature is organized in levels, and the pattern at each level is most clearly discerned by abstracting from the details of the levels far below”
Example: Consciousness
Psychological Measurement
Crosses the levels

- Psychology is a very broad field of investigation
  - Hormones, chemicals, genes to ...
  - Politics of nations
Physiological

- Hormones (estrogen), Neurohormones (dopamine) & Neurotransmitters (Serotonin)
- Genes
- Circadian cycles
- PET Scans
- CAT scans
- EEG (electroencephalograph)
- fMRI
- Pupilometer
Based substantially on Donder’s subtraction method
Interviews and Judgment

- Job Interviews
  - Structured vs. Unstructured
- Clinical Interviews
- Observer Attributions
- Jury deliberation

- Often categorical
- Dawes – Mechanical vs. judgmental accuracy
Ability & Achievement

- Multiple Choice
- Constructed Response
- Physical Abilities
  - A tale of Firefighters and Ice skaters...
Self-Report

- Personality
- Experiences & Biodata
- Attitudes
- Beliefs
- Affect?

- Cognitive process?
  - Nisbett & Wilson (1977). Telling more than we can know.
Other Report

- Peers
- Significant others
- Family
- Teachers
- Supervisors

- Opportunity to observe slices of the person
- Attribution problems
- The “other” also has a personality and fallible memory
Observational

- **Narratives**
  - attempts to record as much as possible of what happens within the focus of the observation.
  - open-ended, flexible & Rich
  - time consuming to both record and interpret
  
  - **Diary description**
    - a chronological record of behavior made after the behavior occurs
  
  - **Anecdotal record**
    - a descriptive narrative, recorded after the behavior occurs
  
  - **Running record**
    - a sequential record over a given time recorded while the behavior is occurring
  
  - **Specimen description**
    - detailed notes on an identified situation, recorded while the behavior is occurring
Observational

- **Time Sampling**
  - an observation of what happens within a given period of time, coded with tallies or symbols while the behavior is occurring; used to document the frequency of specific behaviors.
    - more objective than narrative records, less time consuming, and it offers a way to observe and record two or more people simultaneously.
    - closed-ended, limited to what happens in the specified time interval, and lacking in behavioral and contextual detail.
Observational

- Event Sampling
  - An observation of an event that has been defined in advance and what happens before and after, recorded briefly while it is taking place
    - Objective
    - Event sampling is closed-ended and limited
Experience Sampling

**Advantages**

- quantitatively study in-depth, everyday experiences, moods, and behaviors in their natural context
- measures administered at random or scheduled intervals during the activity of interest to researchers
- less subject to biases in recall of events, mood, behaviors, or other psychologically important variables
Experience Sampling

- Disadvantages
  - Intrusiveness
  - Missing data & lack of commitment
  - Interruptions
  - Hawthorne effects?
  - Huge amounts of data
Internet Assessment

- Equivalence?
  - Time of day
  - Context
  - Security
  - Breaks or pauses
  - Technological problems – missing data
Measurement Mistakes

- Phrenology
More Mistakes

- Craniometry ...
  - Gould’s Mismeasure of man
  - Broca - demonstrated that women are inferior to men because of their smaller crania. He argued against higher education for women because their small brains couldn't handle the demands
  - Effect of societal values on science
  - Measures are like handguns
    - Nothing wrong with measuring skulls... it’s the inference that must made cautiously and responsibly
More Mistakes

Polygraph
More Mistakes

Handwriting Analysis

I was putting up my new series of syndicated articles on handwriting analysis for newspapers and took the liberty of greatly appreciating

I assume you

Ellis Parker Butler

PLATE 8. Cool, controlled type of personality shown in this writing of author Ellis Parker Butler.
Probable Mistakes

- **Penile Plethysmographology**
  - Started in Czechoslovakia to detect sexual preference
  - Now used in pedophilia cases
  - Major issues with reliability and fakability (validity)
Constructing Multiple Choice and Self-Report Measures

1. Clear Construct Definition
   - What is it…and what isn’t it!
   - Identify & define sub-dimensions
     - goal orientation example

2. Specification Chart
   - Relative importance of sub-dimensions
   - Number & type of items per dimension

3. Identify Population of interest

4. Review existing measures/interview sample

5. Select scaling method

6. Write Items

7. Pilot Test & Item Analysis

8. Item Revision
Item Writing Principles

- See distributed file
- General principles
  1. vocabulary level - should match the population
     - Use reading level algorithms to check
  2. avoid slang - differences in familiarity
  3. avoid outdated/obsolete terms
  4. Don’t use conjunctions
  5. avoid absolute terms - difficult to defend due to exceptions
  6. avoid unfamiliar abbreviations
  7. use non-sexist language – pronouns and possessive verbs for both sexes
  8. brevity – item stem and alternatives brief with all necessary information
  9. one content topic per item - no compound items
  10. Items should cover domain of construct
Writing the Item Stem

1. Don’t include implicit assumptions in the stem (e.g., how concerned are you about…)
2. one clear, central problem completely
3. meaningful without reference to alternative
4. complete enough to serve as a short answer question
5. Avoid irrelevant information
6. avoid awkward, weak sentence structure
7. avoid vague or ambiguous words – more than one interpretation possible
8. avoid negative prefixes and double negatives – increases difficulty
9. emphasize negative terms – prepares mental set (underline)
10. use left-branching for modifying clauses added to the main clause
11. avoid dangling participles
12. avoid passive negative voice
13. avoid central embedding – clause between subject and verb
Constructing Alternatives

1. 4 or 5 alternatives per item – one correct and 3 or 4 distractors

2. All alternatives should be plausible, especially the distractors
   - Answered by those with knowledge correctly and incorrectly by those without knowledge
   - Increases discriminability and reliability of scores
   - Each distractor should be selected by a proportion of the examinees without knowledge
   - Distractors of equal difficulty and attractiveness to correct response

3. One best answer
   - Careful of evidence for incorrect alternatives
   - Obtain opinion of experts
   - Document source of correct answer
Constructing Alternatives

1. “all of the above” alternative – lists of characteristics/alternatives
   - several solutions of equal merit, no obviously wrong distractor
   - all alternatives of equal difficulty
   - proportionate number of times used in items – 50%
   - correct response a proportionate number of times
   - last alternative

2. “none of the above” alternative – for correct answer test (math, spelling)
   - not appropriate for “best answer” tests
   - solutions of equal difficulty an merit
   - used in terms of proportionate number of times – 50%
   - correct response a proportional number of times
   - last alternative
   - lengthens test time

3. do not use “all” and “none” alternatives in the same test
Constructing Alternatives

1. correct answer in same position and equal number of times
2. position of correct response should be randomized
3. numerical sequences should be in non-random order
## Constructing Alternatives

### Direct clue errors

<table>
<thead>
<tr>
<th>distractors</th>
<th>correct answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>specific instances</td>
<td>general rule</td>
</tr>
<tr>
<td>short statements</td>
<td>long statements</td>
</tr>
<tr>
<td>negative statements</td>
<td>positive statements</td>
</tr>
<tr>
<td>positive statements</td>
<td>negative statements</td>
</tr>
<tr>
<td>unqualified statements</td>
<td>qualified statements</td>
</tr>
<tr>
<td>means to end</td>
<td>end product</td>
</tr>
<tr>
<td>professional jargon</td>
<td>basic, simple, concise terms</td>
</tr>
<tr>
<td>laymen language/general</td>
<td>technical definition</td>
</tr>
<tr>
<td>incomplete statement rule</td>
<td>complete statement rule</td>
</tr>
<tr>
<td>all point in common direction</td>
<td>point in opposite</td>
</tr>
<tr>
<td>direction or frame of reference</td>
<td>frame of reference</td>
</tr>
</tbody>
</table>
Constructing Alternatives

More direct clues
- items asked for best plan, combination or sequence
  - one alternative contains all steps
  - one alternative omits a certain step
  - one alternative has a unique set of elements
- obviously wrong distractor – direct opposite of correct answer
- alternative association/clang association – similarity in sound of item stem and correct answer
- duplicating language – similar phrases or synonyms in stem and answer
- items asking the meaning of self descriptive terms – infer answer from common sense
- overlapping alternatives among items – alternatives or answer to one item gives away answer to another item (ex.: same alternative, different answer)
- item duplicity – inadvertently give away answer to other item with extraneous information in stem of alternatives

consequences of direct clue errors
- all people answer the item correctly – no discrimination
- clues used only by knowledgeable testee – less harmful
- item easier for test-wise individual – able to infer correct answer
  - measures test wiseness rather than subject matter, lowers difficulty of test, testees can determine the correct answer through reason rather than knowledge
Constructing Alternatives

- **Indirect Clue Errors** - less obvious, require more common sense reasoning to detect and infer correct answer
  1. Grammatical construction of stem and alternatives are inconsistent except for the correct alternative, grammatical inconsistencies - source of confusion
    - tense of verb
    - use of articles “a”, and “an”
    - use of pronoun without appropriate referent in stem
    - use of plural and singular nouns
  2. Overlapping distractors - say same thing in different ways
    - neither is “one best” answer
  3. Overlapping/logically impossible numerical series
    - cancel overlapping series out
  4. Stray alternatives - not parallel with others alternatives in form or conceptual level
    - ambiguous and multiple right answers
  5. Concept split among alternatives - stem requires selection of appropriate consequence and alternatives from two separate groups
    - do not follow from the central problem in the stem
    - represent further qualifications of stem that are reasononable
  6. Tag along phrases - embellish alternative with implausible phrases to make the distractor more attractive
    - discredits the distractor
Item Response Formats

- Know what your anchors mean to respondents (see Friedman & Amoo (1989))
- Don’t force a response (e.g., 4 point scales)
  - Respondents will express an attitude or belief that they don’t actually hold
- Scale should be balanced (equal number of good and bad scale points)
- Understand that the asymmetry of comparisons (do you prefer $x$ to $y$ is not the same as $y$ to $x$)
- # of scale points = 7+ or -2 (5-9)
- Much more on this when we get to scaling