Observation & Measurement II:
Observational Methods
Interobserver Agreement
Treatment Fidelity

Class 6
SPCD 619

Operational Definitions

☐ Are objective – refer to observable components of the behavior
☐ Are clear and unambiguous
☐ Require little or no inference
☐ Include
   ☐ A descriptive name
   ☐ General definition
   ☐ Elaboration that describes the critical parts of the behavior
   ☐ Typical examples of the behavior
   ☐ Near nonexamples of the behavior

Target Behavior and Data Collection

“Hair pulling was defined as touching the fingers to the scalp, eyebrow, or eyelashes for all 3 participants.”
Coding Manuals

- “a set of rules, definitions, examples, and near nonexamples that guide the observers in counting and/or indicating the duration of the behaviors of interest” (p. 36, Yoder & Symons, 2010)

Ways to find these

- Obtain from authors of previous studies
- Write your own using definitions from the existing literature or definitions developed from experts

Coding Manuals should contain:

Behaviors
- Operational definitions
- 3 examples and near nonexamples
- Symbols/codes used
- Log of changes that you make as you change/clarify/update system

Coding Procedures
- General guidelines
- What to do to prepare for observation
- “...” during the observation
- “...” after the observation
- Sample data forms or screen shots of computerized systems

Reliability
- Plan for conducting IOA checks
- Plan for conducting procedural reliability checks

Data Collection

1. Pose an experimental question
2. Identify and define relevant behaviors
3. Select behavioral dimension(s)
4. Select your measurement system
   - Consider the observational settings
   - Consider the advantages and disadvantages
5. Identify two or more people to conduct the observations

Table 1
Quality Indicators Within Single-Subject Research

<table>
<thead>
<tr>
<th>Dependent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variables are described with operational precision.</td>
</tr>
<tr>
<td>Each dependent variable is measured with a procedure that generates a quantifiable index.</td>
</tr>
<tr>
<td>Measurement of the dependent variable is valid and described with replicable precision.</td>
</tr>
<tr>
<td>Dependent variables are measured repeatedly over time.</td>
</tr>
<tr>
<td>Data are collected on the reliability or interobserver agreement associated with each dependent variable; and</td>
</tr>
<tr>
<td>IOA levels meet minimal standards (e.g., IOA = 80%, Kappa = 0.69%).</td>
</tr>
</tbody>
</table>
Selecting a Measurement System (Yoder & Symons, 2010)

- **System has**
  - Behavior sampling method
  - Participant sampling method
  - Coding decision recording method

**Behavior Sampling**

- **Continuous**
  - Code entire observation session
  - Most complete but most expensive method
  - Can use a computer system to measure onset/offset (timed event) or use a simple tally system (event recording)
- **Intermittent**
- **Interval**

**Event Recording**

- **Advantages:**
  - Useful across many topographies of discrete behaviors (movement cycles)
  - Easy to use – most direct and accurate measure of the number of times a behavior occurs
- **Disadvantages**
  - Requires continuous observation
  - Can’t be used with continuous (i.e., non-discrete) data
  - Inaccurate with high frequency behaviors
  - Confounds frequency with duration of responding
Behavior Sampling

- Continuous
- **Intermittent**
  - “periodic intervals are observed and all instances of key behavior occurring in that interval are recorded” (Yoder & Symons, 2010, p. 55)
  - Cheaper method; can do timed event or event sampling; not used as much as other methods
- Interval

**Interval**

- (estimates duration/frequency)
  - Divides observation into discrete intervals and records whether a response occurred...
    - throughout an interval (whole interval)
    - during a fixed time period at the end of an interval (momentary time sampling)
    - during any time in an interval (partial interval)
- Is commonly used

---

**Whole Interval Recording**

Does the behavior occur **throughout** the interval? (i.e., during the entire interval)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Behavior: [ ] [ ] [ ] [ ] [ ] [ ] [X] [ ] [ ] [ ]

10%
Whole Interval Recording

Student: Alice  
Start time: 8:05  
End time: 8:20  
Setting: morning recess

Definition of behavior:
- Social Interaction: within 2 feet of another child with interactions
- Parallel Play: within 2 feet of another child, no interactions
- Alone: further than 2 feet from anyone
- Organized Play: interacting with others in an organized game or activity
- No Code: none of the above

Observer: Mr. Fables (paraprofessional)

<table>
<thead>
<tr>
<th>Time</th>
<th>Social Interaction</th>
<th>Parallel Play</th>
<th>Alone</th>
<th>Organized Play</th>
<th>No Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>00&quot;</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>o</td>
</tr>
<tr>
<td>10&quot;</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20&quot;</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30&quot;</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40&quot;</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50&quot;</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60&quot;</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70&quot;</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key: 10" = 10 seconds.

Advantages
- Provides an estimate of behavior that occurs continuously
- Can be converted to percentage

Disadvantages
- Tends to underestimate overall occurrence of responding
- Requires continuous observation (fatigue, time)

Momentary Interval Recording

Does the behavior occur during a fixed period at the end of the interval?

Examples:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50%</td>
</tr>
</tbody>
</table>
Momentary Interval Recording

Advantages
- Very efficient for observers
- Can observe multiple students
- Does not require continuous observation

Disadvantages
- Can underestimate responding
- Requires some type of cueing device

Partial Interval Recording

Does the behavior occur at any time during the interval?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

80%

Behavior: 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1

Partial Interval Recording

Student: John
Start time: 8:05  End time: 8:15  Setting: Seventh-grade English class

Definition of behavior: On-task behavior is defined as appropriately writing, reading, talking about the assignment, or waiting to ask the teacher a question regarding the assignment.

Observer: Ms. Geralds (recess monitor)

<table>
<thead>
<tr>
<th>minute</th>
<th>10&quot;</th>
<th>20&quot;</th>
<th>30&quot;</th>
<th>40&quot;</th>
<th>50&quot;</th>
<th>60&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>5</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>7</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>8</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>9</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>10</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>O</td>
<td>X</td>
<td>O</td>
</tr>
</tbody>
</table>

Key: O = Behavior was not observed at all during the 10-second interval.
X = Behavior was observed at least once during the 10-second interval.
10" = 10 seconds.

Advantages
- Provides an estimate of behaviors that occur for a brief period of time
- Can be used with high frequency behaviors

Disadvantages
- Tends to underestimate the occurrence of the behavior if larger intervals are used or overestimate if behavior occurs very frequently
- Requires some type of cueing device
Considerations
(from Yoder & Symons, 2010)

If the behavior of interest
- Is of short duration – consider counting number (frequency) (partial interval)
- “represents a state (e.g., attention)” or has a long duration, consider duration (continuous or momentary interval)

Participant Sampling
(from Yoder & Symons, 2010)
- Who do you code if you have more than one participant?
  - Focal – code one participant for a designated time, then another, then another, etc.
  - Multiple pass – code on only 1 participant for the entire session; frequently record session and make multiple passes, coding 1 participant at a time
  - Conspicuous Sampling – observe the group and record on individuals engaged in “predefined conspicuous behaviors”

Coding Decision Recording Method
- Direct, real-time observation
- Recording for later coding (audio or video)
- Paper/Pencil
- Computer-based program
Where & When to Observe

- Consider relation of setting to the research question
- Observations should reflect what occurs in the settings (events)
  - Selective sampling

Choosing and Training Observers

- Choose individuals who have sufficient time to complete training and are dependable.
- Provide sufficient training prior to beginning study
  - Overview of study (but may want naive observers)
  - “how tos” of data collection
  - Copy of the code book
  - Practice, practice, practice (adjust coding rules as new considerations arise)
  - Establish a minimum criterion for IOA before proceeding with data collection (80 – 90%)
  - Retrain periodically
  - Show your appreciation

Reliability

- Consistency of measurement across conditions, regardless of who is the observer
- Reliability is NOT a measure of the “truth” of the behavior! It is a measure of consistency
  - Interobserver Reliability or Agreement (IOA)
    - A second observer independently measures the target behavior(s). The two results are compared for accuracy, yielding a coefficient or percentage of agreement.

Table

<table>
<thead>
<tr>
<th>Quality Indicators Within Single-Subject Research</th>
<th>Horner et al., 2005, pl. 174</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variable</strong></td>
<td></td>
</tr>
<tr>
<td>• Dependent variables are described with operational precision.</td>
<td></td>
</tr>
<tr>
<td>• Each dependent variable is measured with a procedure that generates a quantifiable index.</td>
<td></td>
</tr>
<tr>
<td>• Measurement of the dependent variable is valid and described with replicable precision.</td>
<td></td>
</tr>
<tr>
<td>• Dependent variables are measured repeatedly over time.</td>
<td></td>
</tr>
<tr>
<td>• Data are collected on the reliability or interobserver agreement associated with each dependent variable, and IOA levels meet minimal standards (e.g., IOA &gt; 80%; Kappa &gt; 0.6).</td>
<td></td>
</tr>
</tbody>
</table>


IOA: Total Percent of Agreement
(frequency-ratio approach)

\[
\text{smaller number} \times 100\% = \text{percent of agreement}
\]

Observer | Number of Times Student Raises Hand | Total Recorded
--- | --- | ---
Observer 1 | x | x | x | x | x | 5
Observer 2 | x | x | x | x | 4
\[
\frac{4}{5} \times 100\% = 80\% \text{ agreement}
\]

IOA: Interval Agreement
(combined, point-by-point)

\[
\text{agreements} \times 100\% = \text{percent of agreement}
\]

Observer | Number of Times Student Raises Hand | Total Recorded
--- | --- | ---
Observer 1 | x | x | x | x | Total = 5
Observer 2 | x | x | x | Total = 3
\[
\frac{3}{4 + 3} \times 100\% = 43\% \text{ agreement}
\]

IOA: Interval Agreement
(combined, point-by-point)

\[
\text{agreements} \times 100\% = \text{percent of agreement}
\]

Observer | Number of Times Student Raises Hand | Total Recorded
--- | --- | ---
Observer 1 | x | Total = 1
Observer 2 | x | Total = 0
\[
\frac{6}{1 + 6} \times 100\% = 86\% \text{ agreement}
\]

IOA: Occurrence/Nonoccurrence Agreement

1. **Agreement of Occurrence**
   \[
   \text{Agreement} \times 100\% = \text{percent of agreement}
   \]

2. **Agreement of Nonoccurrence**
   \[
   \text{Agreement} \times 100\% = \text{percent of agreement}
   \]

   Report as two separate statistics.
How Frequently to Conduct Reliability Checks

- Collect IOA in a minimum of 20 - 33% of observational sessions per condition, for each participant for each dependent variable
  - Use higher percentage if situation calls for this

Procedural Reliability (sometimes also called Treatment Integrity or Fidelity)

- Measure of how the closely the intervention was implemented according to plan

- Requires
  - Operational definition of the intervention (IV)
  - Decision of what aspect(s) of intervention to measure
  - Training of observers to conduct assessment of intervention implementation

THE TREATMENT OF SEVERE BEHAVIOR PROBLEMS IN SCHOOL SETTINGS USING A TECHNICAL ASSISTANCE MODEL
Northup et al., 1994)

Treatment integrity. The five teacher responses to student target behaviors were scored to assess the integrity of all assessment and intervention procedures. Each assessment and intervention procedure always specified a particular consequence that was to follow any occurrence of target behavior immediately. The percentage of student target behaviors that were followed (within two 10-s intervals) by the specified consequence served as an index of treatment integrity and provided a measure of the accuracy with which assessment and intervention procedures were implemented. Treatment integrity was scored only when the behavior for which a consequence was prescribed occurred.

Thinking Ahead on the Research Proposal:
At this point you should

- Have identified a general topic
- Begun locating and reading relevant literature so that you can
  - More clearly focus the specific research question you are asking
  - Begin to define possible targets (DV) and develop possible interventions (IV)
Next Week

- Discuss and practice visual analysis
  - Do the readings so you can participate in class activities

- Bring a laptop and flash drive to class