Inter-Rater Reliability Training Exercises for CATI Supervisors

Lisa Klein
lklein@ssc.wisc.edu
(608) 265-5463
Importance of Live Monitoring

- Live monitoring...
  - ensures compliance with client expectations
  - provides an opportunity for interviewer performance feedback
  - provides a forum for interviewer-rater calibration of expectations
Importance of Inter-Rater Reliability

• **Inter-Rater reliability…**
  - ensures fairness of standards
  - maintains clarity between raters
  - facilitates accuracy of data collection
Challenges of Inter-Rater Reliability Training

- Balancing competing needs
  - Trainings should be **thorough, instructive, comprehensive**
  - Training should **meet the needs of individual raters**
  - Trainings should be as **cost-effective and time-efficient** as possible
Training Goals

- Achieve greater reliability between raters by **calibrating** to a norm
- **Increase scoring consistency** between evaluation categories
- Arrive at **consensus on definitions** and expectations
- Calibrate raters to **supervisor** expectations
- **Forum to discuss issues** related to evaluation
- Present examples of complex issues
- Discuss challenges of interview variability
Inter-Rater Reliability Training Exercises for Supervisors

Current UWSC Data Quality Evaluation Measures
Evaluation Measures- Types of Evaluations

• “Full” monitorings
  - conducted by staff of 15 raters
  - occur at least once per month
  - complete interview evaluation
  - review of calling efficiency
  - extensive discussion with interviewer
  - extensive performance documentation
Evaluation Measures- Types of Evaluations

- "Random" monitorings
  - take place constantly
  - small selection of interview
  - limited interviewer follow-up
  - limited performance documentation
Evaluation Measures- Technology

- Silent-entry monitoring system
- VNC software
- Web interface
- All-call recording capabilities
- In-house audio server
Evaluation Measures - Evaluation Criteria

• 6 primary criteria
  - Reads question verbatim
  - Reads entire question when R interrupts
  - Probes appropriately for codeable response
  - Does not interpret for R
  - Remains objective
  - Proper use of verification

• 5 point evaluation scale (Excellent to Unacceptable)
• Universally agreed upon definitions and descriptions
Evaluation Measures- Evaluation Criteria

- Other evaluation criteria
  - ability to answer questions about survey
  - appropriate refusal conversion efforts
  - records open ended questions verbatim
  - codes close-ended responses correctly
  - appropriate pace and clear speaking voice
  - trains respondent
  - keeps R “on track”
  - professionalism
Inter-Rater Reliability Training Exercises for Supervisors

Current Data Quality Training Processes
Current Inter-Rater Reliability Training Methods

• **Group Monitoring**
  
  - rater staff *listens to and watches* a recorded full monitoring together (large group setting)
  
  - supervisor of raters selects interview and oversees training
  
  - VideoLAN playback technology (spliced to audio file)
  
  - raters score independently
  
  - group discussion of each category to calibrate scoring criteria and individual expectations
Group Monitoring - Advantages and Challenges

• Advantages
  - entire group meets to discuss criteria
  - conducted with a supervisor (calibration of supervisor-rater expectations)
  - cost-effective
  - supervisor chooses recordings

• Challenges
  - group dynamics can inhibit participation of some
  - single exercise takes several hours (depth, but not breadth)
Current Inter-Rater Reliability Training Methods

- Dual Monitoring
  - Raters listen to a series of live interviews in pairs utilizing a rotation system
  - Silent entry monitoring using VNC technology
  - Raters score independently
  - Raters discuss evaluation before giving final score to interviewer
Dual Monitoring- Advantages and Challenges

• Advantages
  - live monitoring creates a more realistic environment
  - raters work one-on-one with each other
  - opportunity to hear wide range of interviews

• Challenges
  - more time-intensive and cost-intensive
  - calibration within pair, but not necessarily group
  - lack of group decision-making abilities
  - variability of live interviews
Introduction of a New Calibration Tool

- Benchmark Monitorings
  - Indicate areas of focus in training
  - Two “benchmarks” conducted; one prior to all training interventions, and one following
  - New RecVNC technology allows live interviews to be recorded, then played back with VideoLAN software, so playback includes sound and sight of live interview
  - Listen independently, score independently, and do not discuss until after all have completed
Inter-Rater Reliability Training Exercises for Supervisors

A New Approach: Balancing Training Methods
Balancing Training Methods

• Inclusion of multiple training techniques
  - Quarterly structure
    1. Initial benchmark exercise to indicate areas of focus
    2. Group monitoring exercise to reach consensus, provide opportunities for supervisor guidance and structure
    3. Dual monitoring to provide variability, in-depth focus, realistic decision-making
    4. Second benchmark exercise to track reliability increases
Balancing Training Methods

• Does the new model meet training goals?
  ✓ Calibration between raters to an established norm
  ✓ Increased scoring consistency (via benchmarks)
  ✓ Consensus between raters on definitions
  ✓ Calibrate raters to supervisor expectations
  ✓ Forum to discuss issues and make group decisions
  ✓ Opportunity for supervisor to present complex examples
  ✓ Discuss challenges of interview variability and coding
Qualitative Supervisory Feedback

- Strong support of new model
- Benchmarks perceived as useful, informative
- Strong self-regulation in terms of following activity guidelines
- Preference for splitting time between training activities
Technology Info

- VideoLAN: [http://www.videolan.org/vlc/](http://www.videolan.org/vlc/)

*All programs listed above have free software downloads available.*
Inter-Rater Reliability Training Exercises for CATI Supervisors

Questions?

Lisa Klein
lklein@ssc.wisc.edu
(608) 265-5463