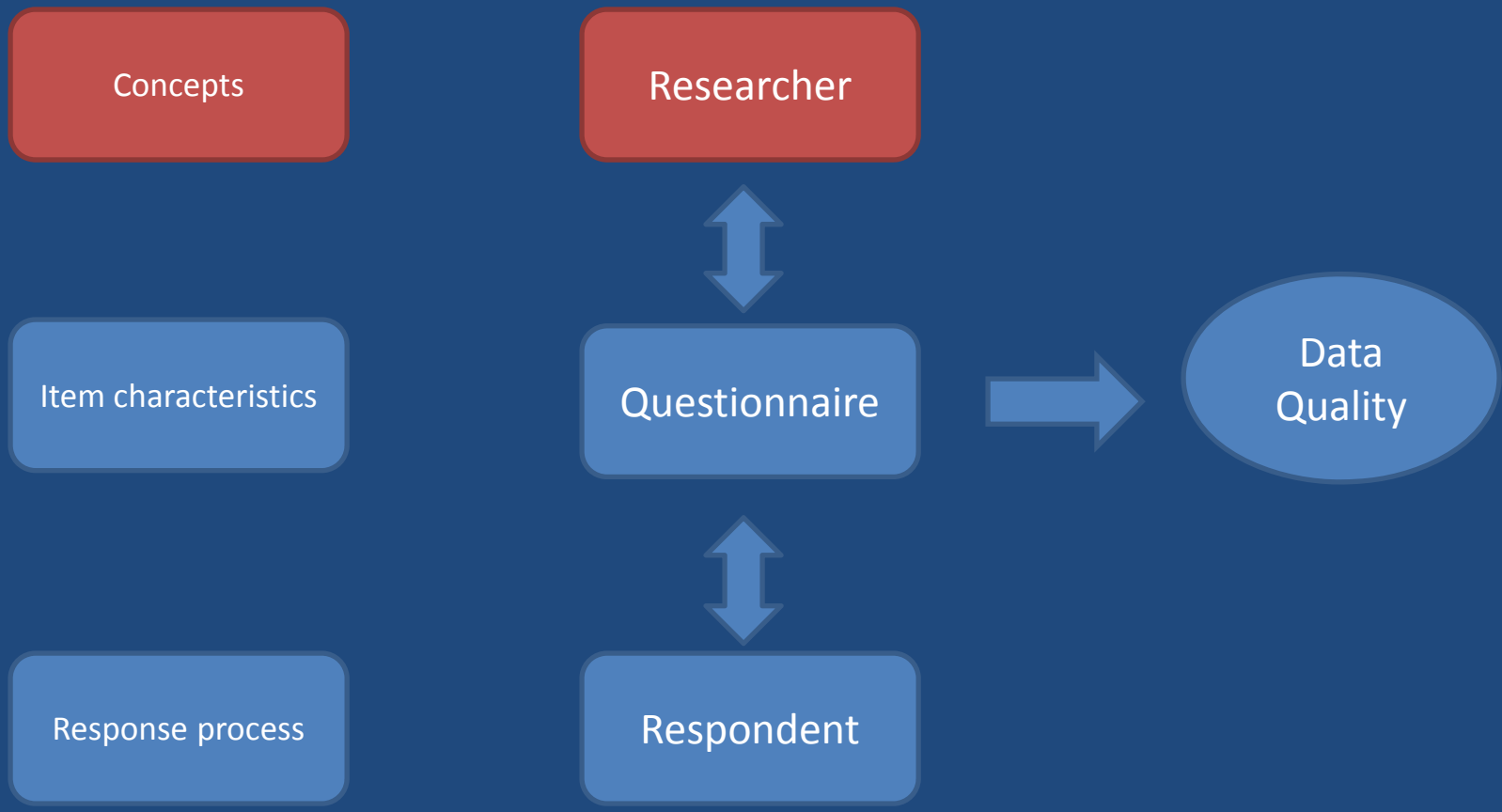
A Comparison Of Ex-ante, Laboratory, and Field Methods for Evaluating Survey Questions

Aaron Maitland, Westat

Background and Motivation



Question Evaluation Methods

- **Online computer tools**
  - Question Understanding Aid (QUAID): Online tool that analyzes questions for various linguistic characteristics that may be problematic.
  - Survey Quality Prediction Tool (SQP): Computer tool used to code characteristics of a survey questions and predict data quality.
- **Expert methods**
  - Expert Review: Questionnaire reviewed independently by individual experts.
  - Questionnaire Appraisal System (QAS): Standardized form to check questions for specific problems.
- **Laboratory methods**
  - Cognitive interviewing: Subjects think aloud or cognitive probing techniques are used to understand responses to the questions.
- **Field methods**
  - Behavior coding: Code the interviews for problematic interviewer or respondent behavior.
  - Response latency: Measure time between the end of the reading of a question and beginning of the answer.
- **Reliability and validity**
  - Test-retest reliability
  - Record check

Previous Research

- Study designs
  - Descriptive
    - Describe problems found by different methods
    - No inferences made
  - Exploratory
    - Compare methods on number and type of problems
    - Agreement or correlational statistics are also common
  - Confirmatory
    - Predicts quality of the data in the field
    - Dependent variables are either indirect assessments of quality in the field such as behavior coding results or direct assessments such as reliability or validity

Previous Research

- Studies have produced inconsistent findings (Presser and Blair, 1994; Rothgeb, Willis, and Forsyth, 2001; Willis, Schechter, and Whitaker, 1999; Yan, Kreuter, and Tourangeau, 2012).
- Very little research on how well pretest methods predict data quality (Dykema, Lepkowski, and Blixt, 1997; Hess, Singer, and Bushery, 1999; Yan, Kreuter, and Tourangeau, 2012).
- When data quality is studied, usually only one method is studied (Dykema, Lepkowski, and Blixt, 1997; Hess, Singer, and Bushery, 1999).

Outstanding Issues

- What are the circumstances in which the methods agree?
- How can we package the methods together to identify problems that affect data quality?
- What is the relative effectiveness of the methods?

Hypotheses

- Complementary methods hypothesis
  - Using multiple methods to predict problems or reliability is better than using a single method.
- Test environment hypothesis
  - Methods implemented in a more realistic setting should be most closely related to data quality.

Data Sources

- Primary data come from the 2006 JPSM survey practicum.
  - Questionnaires included 88 factual and attitudinal questions.
  - Interviews for the final study were recorded.
  - Study included a reinterview of 53 questions two weeks after the initial interview.
- Supplementary questions from record check studies from the literature.

Data Preparation: Qualitative Coding for Ex-ante and Laboratory Methods

- Semantic I: Problems with question structure
  - Information overload, sentence structure, transition problem
- Semantic II: Problems with meaning of terms
  - Vagueness or ambiguity, technical or common terms not understood, double-barreled
- Respondent task I: Problems with recall
  - Recall is difficult, response is resisted
- Respondent task II: Problems with response categories
  - Overlapping or insufficient response categories, response categories not appropriate
- Respondent task III: Problems with sensitivity
  - Social desirability
- Other problems
  - Procedural issues, analysis issues

Data Preparation: Field Methods

- Behavior coding
  - Interviewer codes
    - Exact reading, slight change, major change, probing
  - Respondent codes
    - Adequate answers, qualified answers, inadequate answers, don't know, refusal, respondent interruption, pauses or fillers
- Response latency:
  - Used recordings to measure time from the end of the reading of the question by the interviewer to the onset of the answer by the respondent

Prediction of the Accuracy of Questions

Predictor	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	.87**(.02)	.88**(.02)	.88**(.02)	.85**(.03)	.81**(.02)
Expert review	-.10**(.02)	-.09**(.02)	-.09**(.02)		
QAS	.02(.02)			-.05**(.02)	
Cognitive interviewing	-.28**(.12)	-.24**(.11)			-.26*(.14)
R-squared	.41	.41	.36	.10	.07
**p<.05, *p<.10					

Prediction of Field Results

- Dependent variables: behavior coding, response latency, item nonresponse
- Independent variables: results from QUAID, SQP, expert review, QAS, and cognitive interviewing
- Complementary methods hypothesis:
  - Compare full and reduced models. Can the model be reduced to a single method?
- Test environment hypothesis – Methods implemented in a more realistic setting should be most closely related to data quality.
  - Compare model fit across models (AIC)

Prediction of the Level of Problems (Multilevel Models)

Hypothesis	Adequate Answers	Requests for Clarification	Item Nonresponse	Response Latency
Complementary methods	Supported	Supported	Supported	Not Supported
Predictors in reduced model	QAS sensitivity problems and Cognitive interviewing recall problems	Expert Review, QAS, Cognitive Interviewing problems with question meaning	QAS response category problems and Cognitive interviewing recall problems	QAS recall problems
Test environment	Partially supported	Not supported	Supported	Not supported
ordering of methods	CI and QAS > ER and SQP	CI and ER and QAS > QUAID	CI > QAS > ER > SQP and QUAID	QAS > CI and ER and QUAID > SQP

Prediction of Reliability

- Reinterview occurred two weeks after first interview
- Question level models predicted the index of inconsistency using all available methods.
- SQP total quality (for subjective questions only), QAS recall problems, cognitive interviewing recall problems, QUAID response category problems, and expert review sensitivity issues gave best predictions of reliability.

Conclusions

- Usually necessary to use multiple methods to get the best predictions.
- Direct observation of response process is not always necessary to detect problems with questions, however, cognitive interviewing is often among the best predictors of poorly performing questions.
- Further research needed to understand how to fit the results from online computer tools with traditional method evaluations.
- Future research should be focused on evaluating multiple methods using questions with known psychometric properties.