Some Thoughts on the Use of Field Tests to Evaluate Survey Questionnaires

James L. Esposito
OEUS: Statistical Methods Division
QEM Workshop at the NCHS
21 October 2009

Presentation Objectives

- To share some thoughts on the use of **field-test methodology** to evaluate survey questionnaires, and to do so from the perspective of a survey practitioner
- To provide a conceptual framework that may prove useful/helpful in situating field tests (and other QEM methods) within the broader context of the questionnaire design-and-evaluation process

125 YEARS
BLS

Presentation Outline

- Address Basic Questions about Field Tests
- Field Test Variants and Resources
- Case Study: Displaced Worker Supplement
  - Overview of Methods Used
  - Key Supplement Items: SD1 and SD2
  - Field Test Summaries: 1996, 1998, 2000
- Closing Remarks
- Conceptual Frameworks (discussion phase)

125 YEARS BLS

What are Field Tests?

- **Field tests** are:
  - (usually) complex, collaborative, and resource-intensive evaluation efforts
  - that draw upon the specialized knowledge and skills of individuals and groups of individuals
  - to optimize questionnaire design
  - for the purpose of gathering high-quality data
  - about a particular domain-of-interest (e.g., labor force status; disability; energy use; health status)

125 YEARS BLS

Why Are Field Tests Conducted?

- To identify the principal sources of measurement error in a given questionnaire and to inform subsequent design/redesign decisions to minimize those sources of error

- **Measurement error**: "... a departure from the true value of the measurement as applied to a sample unit and the value provided."
  - Groves, Fowler, Couper, Lepkowski, Singer and Tourangeau, 2004 (pp. 51-52)

125 YEARS BLS

When Are Field Tests Typically Conducted?

- Prior to the Production Phase:
  - When a prototype (or a redesigned) questionnaire has been drafted but not formally evaluated in a field setting
- After the Production Phase:
  - At some point after a survey questionnaire has been fielded, usually to assess data-quality issues or concerns

125 YEARS BLS

Who Are the Principal Field-Test Collaborators?

- Content Specialists
  - subject-matter experts (e.g., sponsors; program managers; academic researchers)
- Design (and Evaluation) Specialists
  - questionnaire and mode(s)
- Interviewers
  - managed by Field Operations Unit
- Respondents

125 YEARS BLS

Where and How Are Field Tests Conducted?

- Where: Ideally, in a natural field setting that closely simulates actual field conditions, and ...
- How: … using multiple evaluation methods in the context of an efficient action plan and timeline

- **Why multiple methods?**
  - Different methods capture/reveal the perspectives and behavior of the various field-test collaborators.
  - All evaluation methods possess strengths and weaknesses. We assume that the weaknesses of any one method will be offset by strengths of the others. [**Table 13**, p. 13]

125 YEARS BLS

Field Test Variants

- Field tests come in “various colors and sizes”
  - from large-scale, multiple-method, multiple-phase undertakings (e.g., redesign of the Current Population Survey)
  - to small-scale, rapid-turn-around pilot tests of questionnaires that gather data on a specific topic
  - and everything in-between (e.g., redesign of DWS and the American Community Survey)

125 YEARS BLS

Necessary Resources

- TIME AND FUNDING: Require sufficient amounts to support and execute the various phases of the design-and-evaluation process, including field tests.
- EXPERIENCED STAFF: The professionals available to make contributions to the process (e.g., content specialists; design-and-evaluation specialists; programmers and authors; interviewers; operations specialists).
- DOMAIN-RELEVANT KNOWLEDGE, INFORMATION AND DATA: The relevant "who, what, what, when, how and why" associated with the domain-of-interest (e.g., health; labor force status; energy use; crime; education).

125 YEARS BLS

Case Study: The Displaced Worker Supplement [DWS]

- **Purpose:** To gather data on the number of persons displaced from jobs over a three-year reference period and their success at finding new employment
- DWS was originally intended as one-time supplement to the CPS (1984), but has been administered every two years thereafter
- Much has changed since the early 1980s

125 YEARS BLS

Working Definition

- "... the term [**displaced worker**] is generally applied to persons who have lost jobs in which they had a considerable investment in terms of tenure and skill development and for whom the prospects of reemployment in similar jobs are rather dim ...
  - (Flaim and Sehgal, 1985, p.4)."
- **Three field tests:** 1996, 1998 and 2000
  - Evaluation methods: Behavior coding; interviewer debriefing; and respondent debriefing
  - Collaborative work: BLS and Census Bureau

125 YEARS BLS

Methods: Behavior Coding

- **Interviewer codes (6)**: Exact reading; minor change; major change; probe; verify and feedback
- **Respondent codes (8)**: Adequate answer; inadequate answer; request for clarification; interruption; DK; REF; and “other”
- Details:
  - Coding was conducted while interviews were in progress using paper-and pencil coding form
  - Multiple exchanges between interviewers and respondents were coded, but analysis focused on the first exchange

125 YEARS BLS

Methods: Interviewer Debriefings

- Focus groups
  - Moderator makes use of a protocol of scripted probe questions [**Table 6,** p. 4, for examples]
  - 10-12 CPS interviewers per FG
- Interviewer logs
  - Written record of problems during interviews
  - Logs enhance retrieval during FG discussions
- Rating form (5-point scale)
  - Useful in quantifying relative magnitude of problems experienced with a given question

125 YEARS BLS

Methods: Respondent Debriefing

- Follow-up probe questions [**Table 10,** p. 6]
  - Used to identify cognitive/conceptual problems that respondents may be experiencing (or be unaware of) when answering specific questions
  - Response-dependent probes developed jointly by content and design specialists
  - If balanced assessments of measurement error are to be undertaken using this method, practitioners need access to relevant **metadata**

125 YEARS BLS

Metadata Defined (1)

- **Metadata:** Any information (verbal or numeric or code, qualitative or quantitative) that provides context for understanding survey-generated data, such as the following:
  - (1) ethnographic observations/information regarding the domain-of-interest;
  - (2) specification of measurement objectives and domain-specific concepts;
  - (3) question wordings, item-specific objectives and ancillary item-specific instructions;

125 YEARS BLS

Metadata Defined (2)

- (4) details regarding data-collection mode(s);
- (5) instructional materials provided to interviewer and/or respondents;
- (6) documentation of prior survey evaluation research; and
- (7) survey-specific classification algorithms and imputation procedures.

125 YEARS
BLS

Displaced Worker Supplement (continued)

- Key DWS Items: SD1 and SD2 [**Table 4,** p.1]
  - Filter (classification) questions

- Relevant metadata [**Table 5,** pp. 2-3]
  - working definition of displaced worker
  - question wordings and specifications
  - definitions of key concepts and terms
  - classification algorithm

125 YEARS BLS

DWS Item SD1

- **SD1.** During the last 3 calendar years, that is January 1995 through December 1997, did you lose a job or leave one because: Your plant or company closed or moved, your position or shift was abolished, insufficient work, *or another similar reason*?
  - <1> Yes (Go to SD2)
  - <2> No (End Displacement Series)

125 YEARS BLS

DWS Item SD2

- **SD2.** Which of these specific reasons describes why you are no longer working at that job?

  [READ IF NECESSARY: If you lost or left more than one job in the last 3 years, refer to the job you had the longest when answering this question and the ones to follow.]

  - <1> Plant or company closed down or moved

    Plant or company still operating but lost or left job because of:

  - <2> Insufficient work
  - <3> Position or shift abolished
  - <4> Seasonal job completed
  - <5> Self-operated business failed
  - <6> Some other reason

[**Note:** Only options 1-3 result in displaced worker classification.]

125 YEARS BLS

1996 Field Test

- Exploratory: Primary focus on SD1 and SD2
- Evaluation methods:
  - BC: Coded 52 “person” interviews, 1 telephone center
  - ID: One FG, 10 interviewers; 1 telephone center
  - RD: Eight follow-up probe questions→false negatives
- Findings [**Table 12,** pp. 10-11]:
  - Evidence of conceptual problems, response problems, design and administration problems
  - Measurement error: Possible undercount of about 25% (false negatives)

125 YEARS BLS

1998 Field Test (1)

- Resource intensive. Focus remained on items SD1 and SD2, but scope of evaluation work expanded.
- Evaluation methods:
  - **BC:** Coded 145 person interviews, 2 telephone centers
  - **ID:** Three FGs, 34 interviewers; 3 telephone centers
  - **RD:** Twenty-two probe questions
- Findings [**Table 12;** also **Tables 7, 8, 9, and 11A-11D**]:
  - Again, evidence of conceptual problems, response problems; design and administration problems
  - Measurement error: False negatives (about 20%); false positives also likely (e.g., temporary jobs; return to old job) but error not quantifiable due to ambiguous specifications

125 YEARS BLS

1998 Field Test (2)

Measurement error decomposition:

- About one-third of false negatives attributed to responses coded as “some other reason” in SD2 (based on verbatim entries):
  - “laid off permanently”; “office closed and had to move”; “bank was bought out so she lost her position”; “program was not refunded” (**Table A-4,** p. 9)
- About two-thirds attributed to inaccurate “no” responses to SD1 (based on respondent debriefing questions and associated verbatim entries)
  - **Tables 11C and 11D**, pp. 7-8, and **Table A-4**, p. 9

125 YEARS BLS

2000 Field Test (1)

- Modest evaluation effort.
- Involuntary job loss (SD1 and SD2) still important, but sponsor interested in expanding supplement to gather data on voluntary job separations
- Evaluation methods [**Table 12**]:
  - BC: Coded 131 person interviews; 2 telephone centers
  - ID: Two FGs, 22 interviewers; 2 telephone centers
  - RD: Eleven probe questions

125 YEARS BLS

2000 Field Test (2)

- Findings: Issues with SD1 and SD2 were similar to those found in 1996 and 1998 (**Table 12**).
  - Measurement error (SD1 and SD2): False negatives (about 29%); false positives likely (temp workers) but displacement concept needs to be more precisely specified.
- Evidence of a somewhat different set of problems for the (debriefing) items gathering data on both voluntary and involuntary job separations
  - Job losers vs. job leavers; early “retirement”
  - Field coding issues (e.g., new item has 20 precodes)
  - Length of reference period (1 vs. 2 vs. 3 years)

125 YEARS BLS

DWS: Current Status

- Not aware of any evaluation work conducted on DWS subsequent to last field test (2000) or of any refinements to the displaced-worker concept.
- DWS due to be administered in 2010 for the three-year reference period, 2007→2009.

125 YEARS BLS

Closing Remarks: Field Tests (1)

- Field tests require collaborative work:
  - Content specialists: Need to know the subject-matter domain and communicate that knowledge to others
  - Design specialists: Need to understand the domain-of-interest and have expertise in questionnaire design-and-evaluation principles and procedures
  - Interviewers: Need to be carefully selected, properly trained, and periodically monitored
  - Respondents: Need to be encouraged to participate and motivated to provide accurate responses (e.g., via use of prudent design features)

125 YEARS BLS

Closing Remarks: Field Tests (2)

Because of what we have learned and think we know about the various phases of questionnaire design-and-evaluation process, survey practitioners have a special responsibility to monitor the functioning of the process and make a determined effort to set in right when it goes off-track.

125 YEARS BLS

Closing Remarks: Q-Bank (1)

- With regard to incorporating field-test research findings within **Q-Bank**:
  - The coding system originally developed for reporting findings from cognitive interviewing appears flexible enough to incorporate findings from multiple-method field tests
  - However, metadata generated from such field tests can be overwhelming and this fact has implications for Q-Bank users and contributors

125 YEARS BLS

Closing Remarks: Q-Bank (2)

- The more evaluation methods employed in any one field test, the more challenging the system becomes for Q-Bank developers, contributors and users alike
- And the more compelling Norman Bradburn's sage counsel regarding successful database systems [ASA 2005]:
  - **simplicity** in system design and use

125 YEARS BLS

Thank you

for attending
this workshop presentation.

125 YEARS
BLS

Situating Field Tests within Broader Conceptual Frameworks

- Survey Lifecycle from a Quality Perspective
  - Groves, Fowler, Couper, Lepkowski, Singer and Tourangeau, 2004 (Figure 2.5, p. 48) [**tan paper stock,** p. 1]
- Parallel paths for measurement and representation
- Focus on the measurement path (left side), specifically the first three boxes, adding a fourth box (observation) prior to the "construct" box:
  - [Observation]
  - Construct
  - Measurement
  - Response

125 YEARS BLS

Expanded Framework (1)

- To better understand how survey data quality is enhanced (i.e., via efforts to minimize measurement error), we will need to expand this measurement lifecycle framework in two directions:
  - **Vertically,** to specifically account for design-and-evaluation phases; and
  - **Horizontally**, to account for the various sources of measurement error

125 YEARS BLS

Expanded Framework (2): Vertical Dimension

- The four elements of the measurement path identified earlier can be viewed as **core design phases** of an expanded questionnaire design-and-evaluation process:
  - P1: Observation ← "observation"
  - P3: Conceptualization ← "construct"
  - P5: Operationalization ← "measurement"
  - P7: (Survey) Administration ← "response"
- And we will also want to incorporate four associated **evaluation phases**
  - P2, P4, P6 and P8 respectively

125 YEARS BLS

Expanded Framework (3): Horizontal Dimension

- One can view the design-and-evaluation process as being subject to five inter-dependent sources of measurement error [adapted from Groves, 1989]:
  - Content specialists
  - Design specialists
  - Interviewers
  - Respondents
  - Mode of data collection

Expanded Framework (4)

- Crossing the two dimensions yields a matrix with 36 uniquely identified cells [c_{ij}] and 4 null cells [**tan pages**, p. 4]
- Each cell represents role- and task-specific activities [cf. Sudman and Bradburn, 1974] specific to a particular phase and error source
- Empty cells [e.g., c_{52}] would indicate that no documentation of activity exists, which could be viewed as problematic
  - design specialist not involved in drafting survey questions

125 YEARS BLS

A Framework Relating Questionnaire Design-and-Evaluation Processes to Sources of Measurement Error: INITIAL DESIGN

			INTERDEPENDENT SOURCES OF MEASUREMENT ERROR (at P7)				
			Questionnaire D-and-E Team		**Information/Data Collection Context**		
		INITIAL DESIGN	*Content Specialist (1)*	*Design Specialist (2)*	*Interviewer (3)*	*Respondent (4)*	*Mode (5)*
Questionnaire *Design*-and-Evaluation Phases	P8	Evaluation	C_{81}	C_{82}	C_{83}	C_{84}	C_{85}
	P7	***Administration***	C_{71}	C_{72}	C_{73}	C_{74}	C_{75}
	P6	Evaluation	C_{61}	C_{62}	C_{63}	C_{64}	C_{65}
	P5	*Operationalization*	C_{51}	C_{52}	C_{53}	C_{54}	C_{55}
	P4	Evaluation	C_{41}	C_{42}	C_{43}	C_{44}	-
	P3	*Conceptualization*	C_{31}	C_{32}	C_{33}	C_{34}	-
	P2	Evaluation	C_{21}	C_{22}	C_{23}	C_{24}	-
	P1	*Observation*	C_{11}	C_{12}	C_{13}	C_{14}	-

Observational base: The **domain-of-interest** as embedded in a "reality" of ceaseless activity (behavior and events) and of durable-yet-mutable relationships (some real, some spurious)—a world within which the observer is an active participant.

Expanded Framework (5)

- Social, cultural and technological change also plays a crucial role in the measurement process
- In the case of panel surveys, moderate-to-rapid change in the target domain can have a substantial effect on the magnitude of measurement error
- Redesign work inevitable in such cases

125 YEARS BLS

A Framework Relating Questionnaire Design-and-Evaluation Processes to Sources of Measurement Error

			INTERDEPENDENT SOURCES OF MEASUREMENT ERROR (at P7 or RP7)				
			Questionnaire D-and-E Team		Information/Data Collection Context		
		REDESIGN	*Content Specialist (1)*	*Design Specialist (2)*	*Interviewer (3)*	*Respondent (4)*	*Mode (5)*
	RP8	Evaluation	C_{R81}	C_{R82}	C_{R83}	C_{R84}	C_{R85}
	RP7	***Administration***	$\mathbf{C_{R71}}$	$\mathbf{C_{R72}}$	$\mathbf{C_{R73}}$	$\mathbf{C_{R74}}$	$\mathbf{C_{R75}}$
Questionnaire	RP6	Evaluation	C_{R61}	C_{R62}	C_{R63}	C_{R64}	C_{R65}
***Redesign* and**	RP5	*Operationalization*	C_{R51}	C_{R52}	C_{R53}	C_{R54}	C_{R55}
Evaluation	RP4	Evaluation	C_{R41}	C_{R42}	C_{R43}	C_{R44}	-
Phases	RP3	*Conceptualization*	C_{R31}	C_{R32}	C_{R33}	C_{R34}	-
	RP2	Evaluation	C_{R21}	C_{R22}	C_{R23}	C_{R24}	-
	RP1	*Observation*	C_{R11}	C_{R12}	C_{R13}	C_{R14}	-
		INITIAL DESIGN	*Content Specialist (1)*	*Design Specialist (2)*	*Interviewer (3)*	*Respondent (4)*	*Mode (5)*
	P8	Evaluation	C_{81}	C_{82}	C_{83}	C_{84}	C_{85}
	P7	***Administration***	$\mathbf{C_{71}}$	$\mathbf{C_{72}}$	$\mathbf{C_{73}}$	$\mathbf{C_{74}}$	$\mathbf{C_{75}}$
Questionnaire	P6	Evaluation	C_{61}	C_{62}	C_{63}	C_{64}	C_{65}
***Design* and**	P5	*Operationalization*	C_{51}	C_{52}	C_{53}	C_{54}	C_{55}
Evaluation	P4	Evaluation	C_{41}	C_{42}	C_{43}	C_{44}	-
Phases	P3	*Conceptualization*	C_{31}	C_{32}	C_{33}	C_{34}	-
	P2	Evaluation	C_{21}	C_{22}	C_{23}	C_{24}	-
	P1	*Observation*	C_{11}	C_{12}	C_{13}	C_{14}	-

Observational base: The **domain-of-interest** as embedded in a "reality" of ceaseless activity (behavior and events) and of durable-yet-mutable relationships (some real, some spurious)—a world within which the observer is an active participant.

Expanded Framework (6)

- The design-and-evaluation process is not necessarily linear (P1→P7):
  - Phases can overlap
  - Movement between phases can be bidirectional and iterative (e.g., only between P1 and P6)
- Work performed inadequately at early phases represent precursors of measurement error at the administration phase

125 YEARS BLS

Questionnaire Evaluation Methods [QEMs]

- Evaluation Phases:
  - Initial Design: P2, P4, P6 and P8
  - Redesign: RP2, RP4, RP6 and RP8
- The optimal choice of a QEM would appear to be phase specific [**tan pages,** pp. 6-7], for example:
  - Participant observation at P2
  - Cognitive interviews and expert panels at P4
  - Questionnaire appraisal systems at P6 (early)
  - Behavior coding, calendar method, focus groups, follow-up probes at P6 and/or P8

125 YEARS BLS

Contact Information

James L. Esposito
Survey Practitioner: Evaluation and Design
OEUS: Statistical Methods Staff
202-691-6368
esposito.jim@bls.gov

BLS
125 YEARS
BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR