

Some doubts about if Cognitive Interviewing findings improve survey estimates? Do Mixed-Methods research

José-Luis Padilla, Ph. D.
Isabel Benítez, Ph. D.
University of Granada, Spain

Presentation at the 2013 QUEST conference, April 9th-11th,
Washington DC

About the authors: José-Luis Padilla, is an Associate Professor of Psychometrics and Research Methodology. Isabel Benítez is Assistant Professor of Research Methodology. Their primary interests are psychometrics, survey methodology, validity, questionnaire adaptation and evaluation, and cognitive interviewing.

Overview

- ✦ Introductory Remarks
- ✦ What doubts do we have about CI and why?
- ✦ Foundations for a broader methodological framework: Validity concept and Mixed research
- ✦ Research case 1: “Trying to identify DIF causes in psychological scales for cross-lingual testing”
- ✦ Research case 2: “Attitudes toward euthanasia: what does the survey questions capture?”

INTRODUCTORY REMARKS

- ✦ Our research program on validity and cognitive interviewing has grown out and focused around the below themes and experiences:
 - ✦ Since the beginning of Survey Research and Psychometrics there have been parallel developments: measurement of attitudes, “five-points scale”, cognitive psychology, etc. (see Groves, 2011).
 - ✦ Current views of Validity in Psychometrics can help survey researchers in dealing with “measurement errors”.
 - ✦ Applied research contracts for Spanish National Statistics Institute, Eurostat, private institutes, etc., and the research on Differential Item Functioning (DIF) in psychological tests and questionnaires for cross-lingual testing.
- ✦ My primary goal for this presentation is twofold: a) to share with you my reflections on how we can resolve doubts about CI usefulness; and b) to comment two research cases we are conducting within MR framework.

WHAT DOUBTS... AND WHY?

- ⊗ Doubts about usefulness of CI are raised as:
 - ⊗ To what extent CI findings do detect “real problems”?
 - ⊗ Are CI findings improving survey estimates?
- ⊗ What are the reasons behind the concerns about CI:
 - ⊗ As Presser et al (2004) suggest “*Questionnaire design and statistical modeling are usually thought of as being worlds apart*” (p. 12). Could this separation be extended to “Questionnaire designers” and “Questions evaluators”?
 - ⊗ Limitations of Total Survey Error framework for dealing with “measurement errors”.
 - ⊗ What does “to improve survey estimates” mean? Does it mean to make “survey estimates” more “accurate”?
 - ⊗ To understand “validity” only as a matter of “survey accuracy”: “*the deviation of a survey estimate from its underlying true parameter value*” (Biemer, 2011; p. 817).

Limitations of TSE framework with measurement errors

- ✦ TSE aims to control all major sources of error in surveys and develop a “metric” not only for quantifying but also optimizing survey design.
- ✦ Major difficulties appear when trying to combine “statistical and non-statistical components” of errors (Groves & Lieberg, 2011):
 - ✦ Inadequate definition of “specification error”: “the concept “implied” by the question differs from the concept that should have been measured” (Biemer, 2011). But: What does “implied” mean? Is it only a matter of “poor communication” between researchers and questionnaire designers?
 - ✦ Is it really possible to include all sources of “measurement errors” in a common metric?
 - ✦ Where ends the “specification error” and begin “measurement errors”? It is not useful to keep a strict distinction between “specification error” and “measurement error”: a) It is not possible to know without resorting to the respondents if the question is capturing the intended concept; and b) all sources of “measurement errors” can undermine the validity of interpretations of the survey estimates for the intended construct.

A BROADER METHODOLOGICAL FRAMEWORK: THE VALIDITY CONCEPT

- ✦ Could survey researchers adopt this validity concept changing “test scores” and “test” with “survey estimates” and “survey”?

“Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests”

(AERA, APA, NCME, 1999)

The validation process implies accumulating evidence for the **proposed interpretations** of the scores.

Validity tenets for survey researchers

- ✦ “Fitness for use” sees survey quality as a complex, multidimensional concept that goes beyond of the TSE paradigm, and include user perspective on survey quality that take “accuracy” for granted and pay extra attention to “questionnaire content”.
- ✦ Validity is not all about “accuracy”.
- ✦ Validity is NOT a property of a question, of a questionnaire, even of survey estimates.
- ✦ Validity is a matter of INTERPRETATIONS. “Interpretations” behind decisions based on survey estimates. So, “to improve survey estimates” should be understood as to support survey estimates interpretations.
- ✦ To validate a survey estimate interpretation requires multiple lines of evidence (e.g., Cognitive interviewing findings).

- ✦ Proponents of mixed research typically adhere to the compatibility thesis as well as to the philosophy of pragmatism:
 - ✦ Pragmatism: Researchers should use the approach or mixture of approaches that works the best in a real world situation.
 - ✦ Integration: Quantitative and qualitative methods can both be used in a single research study.
- ✦ What is a Mixed Research?:
 - ✦ “Research in which the investigator collects and analyses data, integrates the findings, and draws inferences using both qualitative and quantitative approaches or methods in a single study or a program of inquiry” (Tashakkori & Creswell, 2007).

How can MR remove doubts about CI?

- ✦ Conducting CI within a MR in which CI and quantitative methods are integrated through all phases in the question evaluation projects.
- ✦ Planning the MR question evaluation projects according to the best MR design (see Creswell, 1995; for one introduction to MR designs)
- ✦ Examples of MR question evaluation projects:
 - ✦ QDRL project for developing a question about the construct "sexual identity" ahead of their inclusion in the National Health Interview Survey (Miller and Ryan, 2011).
 - ✦ Studies to evaluate a set of disability questions for comparability across the US and six Asian countries. Padilla, Miller, Loeb, and Maitland (2011) combining CI and field tests by "probe questions" and Multidimensional Scaling analyses.
 - ✦ Padilla, Benítez and Castillo (2013, in press) to obtain validity evidence of the APGAR family function scale combining psychometrics and CI findings.

Research case 1: “Trying to anticipate DIF causes in psychological scales for cross-lingual testing”

- ✦ DIF occurs when examinees with the same proficiency level on the characteristic or attribute measured, but who belong to different groups (demographic, linguistic, national or cultural), have a different probability of giving a specific item response” (Millsap & Everson, 1993).
- ✦ Why DIF is a measurement problem? DIF can undermine the validity of cross-national/cultural comparisons.
- ✦ We are interested in DIF in polytomous items: mainly attitude items like PISA Student Questionnaire items.
- ✦ Benitez & Padilla (2013, in press) illustrate how to interpret DIF results (“giving meaning to DIF statistics”), by cognitive interviewing. When comparing US and Spanish students, different interpretation patterns about “terms”, “expression”, “educational and cultural factors”, etc., could explain DIF found in PISA polytomous items intended to measure attitudes toward sciences.

Research case 1: A step forward...

- ✚ In the second phase of the research project, our aim is to examine the usefulness of CI for “anticipating” potential sources of DIF.
- ✚ The research will analyze DIF in PISA 2009 Student Questionnaire focused on: “enjoyment of reading” (Q24); “learning strategies” (Q27); “attitudes toward school” (Q33); and “teacher-students relations” (Q34). In addition, we want to include “cognitive items”.
- ✚ Example of PISA 2009 Student Questionnaire item:
 - ✚ English version: “*Reading is one of my favourite hobbies*”.
 - ✚ *Spanish version: “Leer es una de mis aficiones preferidas”*
- ✚ MR design: QUAN (Multidimensional Scaling) + QUAL (CI) + QUAN (DIF)
- ✚ CI will be conducted in US and Spain to capture differences in interpretation patterns that could predict DIF results.

Research case 2: “Attitudes toward euthanasia: what does the survey questions capture”

- ✦ Several surveys conducted in Spain provided contradictory survey estimates of attitudes toward euthanasia. For example: positive attitudes decrease when questions “go into details”: specific illness or assisted suicide.
- ✦ A MR project is being conducted following the below research design:
 - ✦ QUAL (Focus group) + QUAN (Field Test and Survey) + QUAL (CI).
- ✦ We will conduct 40 cognitive interviews. Caregivers of terminal ill people will be also interviewed.
- ✦ The aim of the question evaluation project is to find out how respondents understand concepts like euthanasia, palliative sedation, assisted suicide, etc., and how they “link” their interpretations to the response categories.
- ✦ We want to link interpretations patterns and evidence of the response processes to survey statistics like “Don’t know” percentages and inconsistent response patterns.

Proposals for QUEST research agenda

- ✦ To contribute to the theoretical background of question evaluation methods
- ✦ To develop mixed methods evaluation models for survey questions
- ✦ To broaden the consensus about theoretical foundations, practices, and... Standards!

REFERENCE

- ✦ AERA, APA, NCME. 1999. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- ✦ Benítez, I., and J. L. Padilla. 2013, in press. Analysis of non-equivalent assessments across different linguistic groups using a Mixed Methods approach: Understanding the causes of Differential Item Functioning by Cognitive Interviewing. *Journal of Mixed Methods Research*.
- ✦ Biemer, P. P. 2011. "Total Survey Error: Design, Implementation, and Evaluation." *Public Opinion Quarterly* 74(5):817–848.
- ✦ Creswell, J. W. 1995. *Research design: Qualitative and quantitative approaches*. Thousand Oaks. CA: Sage.
- ✦ Groves, R. M. 2011. "Three Eras of Survey Research." *Public Opinion Quarterly* 75(5):861–871.
- ✦ Groves, R. M., and L. Lyberg. 2011. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly* 74(5):849–879.
- ✦ Miller, K., and J. Michael Ryan. 2011. *Design, Development and Testing of the NHIS Sexual Identity Question*. Retrieved (http://wwwn.cdc.gov/qbank/report/Miller_NCHS_2011_NHIS_Sexual_Identity.pdf).
- ✦ Padilla, J.L., I. Benítez, I. and M. Castillo. 2013, under review. Obtaining validity evidence by cognitive interviewing to interpret psychometric results. *Methodology*
- ✦ Presser, Stanley et al. 2004. "Methods for Testing and Evaluating Survey Questions." Pp. 1–22 in *Methods for Testing and Evaluating Survey Questionnaires*, edited by Stanley Presser et al. Hoboken, NJ: John Wiley and Sons, Inc.

Thanks for your attention

*Don't hesitate to contact me for comments,
doubts, or suggestions.*

José-Luis Padilla (jpadilla@ugr.es)