RESEARCH SYNTHESIS
THE PRACTICE OF CROSS-CULTURAL COGNITIVE INTERVIEWING

GORDON B. WILLIS*

Abstract   Cross-cultural cognitive interviewing (CCCI) has increasingly been practiced across a range of cultures, languages, and countries, in an effort to establish cross-cultural equivalence of survey questions and other materials, to detect sources of difficulties in answering survey questions for particular subgroups, and to detect problems related to translation from source to target languages. Although descriptions of such studies have proliferated in both the published and unpublished literatures, there has been little effort to reconcile discrepant views, approaches, and findings. The current synthesis reviews 32 CCCI studies located in peer-reviewed journals and books, along with key unpublished sources, to characterize these investigations in terms of their purpose, procedures, and findings. Based on a number of trends in this emergent field, conclusions are made concerning appropriate methods for cognitive testing of cross-cultural instruments, and recommendations are made for future practices that will serve to advance the CCCI field.

Surveys that are variously referred to as cross-cultural, comparative, multilingual, multicultural, and multiregional have proliferated greatly (Harkness, Braun, et al. 2010). Smith (2004) suggests that cross-cultural studies call for special attention to questionnaire development and pretesting, and Beatty and Willis (2007, p. 298) note that "cognitive interviewers increasingly face the need to conduct cross-cultural and multi-lingual testing." Given the extent to which cognitive interviewing is believed to elucidate the underlying manner

in which survey respondents interpret and mentally process survey questions (Willis 2005; Miller et al. 2014), the application of cognitive interviewing to evaluate questionnaires intended for multiple cultures and languages appears to be a natural extension (Chan and Pan 2011; Willis and Miller 2011). A number of studies devoted to cross-cultural cognitive interviewing (CCCI) have appeared in peer-reviewed journals and edited volumes, and even more exist within the unpublished ("gray") literature. The inclusion of overt cultural elements has incorporated sociological and anthropological perspectives into an already interdisciplinary paradigm (Gerber 1999; Miller et al. 2014), and these studies cover a diverse range of countries, cultures, languages, questionnaire types, and cognitive pretesting approaches.

However, it is difficult to judge the efficacy of cognitive testing in meeting the challenges of cross-cultural questionnaire pretesting and evaluation. Fundamentally, cross-cultural applications cannot be assumed to be valid (Goerman and Caspar 2010a, 2010b; Harkness, Edwards, et al. 2010; Park, Sha, and Pan 2013). In particular, if key techniques such as cognitive probing themselves produce differential effects across subgroups, then any resulting variation in behavior may be attributable to artifacts of the measurement process, as opposed to cross-cultural variation in functioning of the survey questions evaluated. For example, it has been argued that Asians are less forthcoming in providing critical opinions (Chan 2010); if so, the appearance that questions are "working" for Asians, relative to a more vocally expressive cultural group, may lead to erroneous conclusions. Extension of cognitive testing to multiple languages also presents significant analysis and interpretation challenges: When bilingual cognitive interviewers are assigned to different language subgroups, it can be difficult to assess whether differential effects across language are due to questionnaire function, as opposed to interviewer effects. Finally, CCCI studies tend to present logistical complexity, and to require careful attention to multiple sequential steps and decision points (summarized in figure 1). To promote further examination of these challenges, the current review assesses what has been learned from CCCI investigations by synthesizing the existing literature, focusing mainly on peer-reviewed publications. I characterize the current state of the science, develop hypotheses concerning specific practices that are effective, and suggest directions for further research to fill gaps or resolve controversies.

Methodology for the Review

Following Johnson (2006), I adopt the general term *cross-cultural* to represent the range of CCCI studies involving cultural and linguistic variation. However, deciding which studies fit within that category is not straightforward, as it could be argued that any testing effort that actively selects respondents (cognitive interview *participants*) from diverse cultural or demographic subgroups (e.g., from varied regions of the US) meets this definition. For current
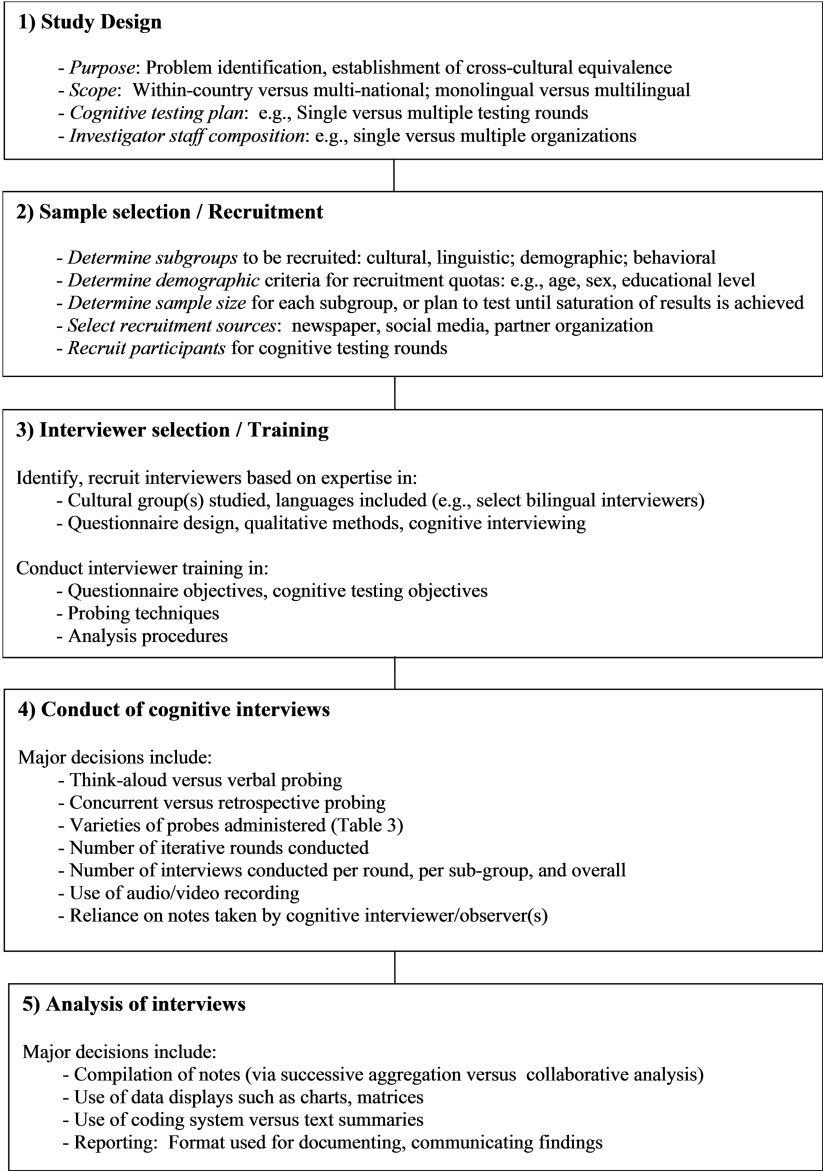
1) Study Design

- *Purpose*: Problem identification, establishment of cross-cultural equivalence
- *Scope*: Within-country versus multi-national; monolingual versus multilingual
- *Cognitive testing plan*: e.g., Single versus multiple testing rounds
- *Investigator staff composition*: e.g., single versus multiple organizations

2) Sample selection / Recruitment

- *Determine subgroups* to be recruited: cultural, linguistic; demographic; behavioral
- *Determine demographic* criteria for recruitment quotas: e.g., age, sex, educational level
- *Determine sample size* for each subgroup, or plan to test until saturation of results is achieved
- *Select recruitment sources*: newspaper, social media, partner organization
- *Recruit participants* for cognitive testing rounds

3) Interviewer selection / Training

Identify, recruit interviewers based on expertise in:
- Cultural group(s) studied, languages included (e.g., select bilingual interviewers)
- Questionnaire design, qualitative methods, cognitive interviewing

Conduct interviewer training in:
- Questionnaire objectives, cognitive testing objectives
- Probing techniques
- Analysis procedures

4) Conduct of cognitive interviews

Major decisions include:
- Think-aloud versus verbal probing
- Concurrent versus retrospective probing
- Varieties of probes administered (Table 3)
- Number of iterative rounds conducted
- Number of interviews conducted per round, per sub-group, and overall
- Use of audio/video recording
- Reliance on notes taken by cognitive interviewer/observer(s)

5) Analysis of interviews

Major decisions include:
- Compilation of notes (via successive aggregation versus collaborative analysis)
- Use of data displays such as charts, matrices
- Use of coding system versus text summaries
- Reporting: Format used for documenting, communicating findings

Figure 1. Key Stages of Cross-Cultural Cognitive Interviewing (CCCI) Studies.

purposes, I limit the review to studies that make use of cognitive interviewing—as defined by key sources (Beatty and Willis 2007; Miller et al. 2014)—in contexts that (a) involve administration of survey questionnaires that are

translated from a source to one or more target languages; or (b) whether or not translation is done, that involve cultural elements differing significantly from that in which the source questionnaire was developed. The latter category includes efforts to evaluate the cognitive testing *process,* when applied to a context other than the Western-based settings where it has typically been conducted—that is, to consider not only "How does cognitive interviewing apply to a questionnaire that has been translated into Chinese?" but also "How well does cognitive interviewing work with Chinese participants?"

To conduct the review, I initially obtained materials from sources likely to contain CCCI studies (without an attempt to exhaustively identify every such study existing in the literature, however):

(a) The contents and reference sections of recent books devoted to cross-cultural survey methods (e.g., Harkness, Braun, et al. 2010); a special issue of *Field Methods* devoted to CCCI (Willis and Miller 2011); and the bibliography of the University of Michigan Cross-Cultural Survey Guidelines (http://ccsg.isr.umich.edu/bibliography.cfm);

(b) An online search of journals containing articles on CCCI (*Bulletin of Sociological Methodology*; *Survey Research Methods*; *Survey Methods*: *Insights from the Field*; *The International Journal of Public Opinion Research*; *Public Opinion Quarterly*; *Field Methods*; *Quality and Quantity*; and *Quality of Life Research*), by first selecting articles containing the terms "cognitive," "pretest," or "interview," then further limiting the search to those devoted specifically to CCCI, and finally checking articles' reference citations for additional eligible publications;

(c) In order to avoid potential effects of publication bias and to represent the extensive unpublished sources containing descriptive details concerning CCCI, searches of the online Proceedings of the American Statistical Association/American Association for Public Opinion Research Annual Meetings) from 1995 to 2013; of the 2014 AAPOR meeting presentations; and of the Q-Bank database of cognitive testing reports (http://wwwn.cdc.gov/qbank).

Analysis of Sources

To focus mainly on empirical sources vetted by scientific review, I then abstracted major elements and procedural details from that subset of studies (a) contained in peer-reviewed sources (academic journals, books); (b) that conducted applications of CCCI (rather than only discussing the topic or covering a single activity such as participant recruitment); (c) that were sufficiently unique (i.e., where multiple publications shared the same data set and the conclusions largely overlapped, only one was selected); and (d) that were sufficiently detailed to provide key items of information. The resulting list of

32 studies, summarized in table 1, represent an eclectic assemblage involving numerous languages, including English, Spanish, Chinese (Mandarin, Cantonese), Korean, Vietnamese, Thai, Bangla, Malaysian, Danish, French, Russian, German, Dutch, Hungarian, Bulgarian, Portuguese, and Maori; and conducted mainly in North America but also in Europe, Asia, Africa, Mexico, New Zealand, India, Bangladesh, and Central and South America. For each study, table 1 lists: (a) the authors; (b) the overall purpose of the study; (c) population subgroups included; (d) languages included; (e) type of survey material evaluated; (f) key procedural features; and (g) major findings and conclusions, including features of CCCI reported by the authors as either useful or problematic. A review of the table elucidated several major issues, distinctions, and conclusions, which I discuss below (where appropriate, I also make reference to unpublished sources).

COMPARISON OF CCCI WITH STANDARD COGNITIVE INTERVIEWING

In assessing the studies within table 1, a fundamental issue is whether CCCI studies depart, either qualitatively or quantitatively, from the "standard" cognitive interview projects described in prior reviews, especially Beatty and Willis (2007) and Willis (2005). Table 1 reveals that one overall objective is very similar to that of standard cognitive interviewing investigations: The conduct of pretesting in order to identify potential respondent difficulties, and more generally, to "repair problems" in tested survey items. Further, most CCCI studies incorporate key procedures that adhere closely to those developed for standard cognitive testing, for example the use of iterative testing involving multiple rounds, the inclusion of both think-aloud and verbal probing techniques, and the application of a variety of probing types. To provide some specificity, table 2 lists illustrative findings from select CCCI studies. Overall, these are very similar to the results of standard cognitive tests conducted over the past thirty years, and further support the contention that cross-cultural cognitive testing is effectively a variant of standard cognitive testing.

Beyond general problem detection, however, an additional feature of CCCI studies is frequently to determine whether the different questionnaire versions illustrate the key property of *cross-cultural equivalence*; that is, whether the range of interpretations associated with the evaluated items varies acceptably between cultural or language groups, given the survey measurement objectives. Observed disparities in interpretation may then be addressed through revision to one or more versions (e.g., the target-language translation). Alternatively, one subgroup's interpretation might not be viewed as more accurate than another, so there may be nothing to "repair" in any tested instrument version. For instance, perceptions of "general health" have been found to differ between Hispanics and non-Hispanics, without either conceptualization being designated as incorrect and in need of modification (Miller et al. 2005). In any event, an increased emphasis on cross-cultural equivalence has led to

Table 1. Summary of Results of 32 Peer-Reviewed Reports Involving the Practice of Cross-Cultural Cognitive Interviewing (CCCI)

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Agans, Deeb-Sossa, and Kalsbeek (2006)	Apply CCCI to evaluate questions for Mexicans	Recent (female) Mexican immigrants (age 18–45)	Spanish (*n* = 10)	Items on women's health (last menstrual period)	Retrospective probing No information on degree of probe flexibility	1) Found CCCI successful for Spanish-speaking Mexican immigrants
Behr, Braun, Kaczmirek, and Bandilla (2014)	Apply embedded (web) probing to evaluate cross-national equivalence	1) Canadians 2) US Americans 3) Danish 4) German 5) Hungarians 6) Spanish	1) English (516) 2) English (524) 3) Danish (537) 4) German (1044) 5) Hungarian (536) 6) Spanish (538) (*n* = 3,695)	A single question on civil disobedience embedded in field survey	Web/embedded probing, with development of descriptive codes from data	1) Conceptualizations of "civil disobedience" varied across countries 2) Web probing useful for obtaining many respondents, but is inflexible, does not allow follow-up probing, and requires significant time for analysis

Continued

Table 1. *Continued*

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Berrigan, Forsyth, Helba, Levin, Norberg, and Willis (2010)	Study item equivalence across recent and long-term Latino immigrants	Hispanics with varied years of US residence; monolingual Spanish, and bilinguals	Spanish (18) English (9) (*n* = 27)	Items on physical activity, acculturation	Single testing round Concurrent probing Flexible probes	1) Many problems were unrelated to translation, but concerned non-linguistic cultural issues and generic problems of questionnaire design 2) No mention of problems associated with CCCI
Chan and Pan (2011)	Evaluate translations brochure accompanying pre-notification letter	English speakers (10), non-English speakers (49) living in US	English (10) Spanish (12) Chinese (12) Korean (13) Russian (12) (*n* = 59)	Advance materials for US Census (multi-lingual brochure)	Retrospective probing Two iterative rounds conducted	1) Chinese and Korean speakers were less survey literate 2) The purpose of a survey needs to be explained to the unacculturated 3) It is unclear whether English and non-English CCCI results are comparable

Table 1. *Continued*

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Daveson, Bechinger-English, Bausewein, Simon, Harding, Higginson, and Gomes (2011)	Assess item equivalence across English and Germans	Residents of England Residents of Germany	English (15) German (15) (*n* = 30)	Scale on attitudes toward end-of-life care (EoLC)	Concurrent and retrospective probing; structured and flexible probes Used grounded, inductive coding scheme	1) Assessed major themes in CI reports to create model of attitudes toward EoLC, combined across countries 2) No mention of problems with CCCI
Fitzgerald, Widdop, Gray, and Collins (2011)	Identify problems with translations and evaluated cross-cultural equivalence, in six-nation study	1) Bulgarians 2) Germans 3) British 4) Portuguese 5) Spanish 6) Swiss	1) Bulgarian (10) 2) German (10) 3) English (29) 4) Portuguese (8) 5) Spanish (18) 6) French (15) (*n* = 90)	Multiple items from the European Social Survey (ESS)	Probing rather than think-aloud Flexible probes	1) Found problems related to translation, cultural portability, source question defects 2) No mention of problems with CCCI
Forsyth, Kudela, Levin, Lawrence, and Willis (2007)	Evaluate CCCI as a means to identify problems with translations	Hispanics, Asians, with US Americans as comparison base	1) Spanish (9) 2) Chinese (9) 3) Korean (9) 4) Vietnamese (14) (*n* = 41)	Tobacco use questions	Two iterative rounds Structured and flexible probes	1) No mention of problems with CCCI 2) Flexible probing is vital 3) Experienced cognitive interviewers are preferable

Continued

Table 1. *Continued*

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Fujishiro, Gong, Baron, Jacobson Jr., DeLaney, Flynn, and Eggerth (2010)	Identify problems with translations to multiple languages	English-, Spanish-, Chinese-speaking home care workers	1) English (10) 2) Spanish (10) 3) Chinese (10) (*n* = 30)	Questions on job-related quality of life	Two iterative testing rounds Concurrent probing	1) No mention of problems with CCCI 2) Nesting of cognitive interviewers within language groups may introduce biases 3) Sample size was insufficient for achieving saturation of results
Goerman and Caspar (2010a)	Identify problems in 2010 US Census questionnaire	Mono- and bilingual Spanish speakers Monolingual English speakers	Round 1: Spanish (44) Round 2: Spanish (41) English (25) (*n* = 110)	2010 US Census questionnaire	Retrospective probing Structured and flexible probes	1) Novice interviewers require standardized probes 2) CI report should be in English with key results in target language 3) Large number of interviews requires coding scheme for analysis

Continued

Table 1. *Continued*

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Goerman and Caspar (2010b)	Compare (a) Spanish-only CCCI with (b) parallel Spanish and English CCCI	Mono- and bilingual Spanish speakers Monolingual English speakers	1) Spanish (85) 2) English (25) (*n* = 110)	2010 US Census questionnaire	Retrospective probing Structured and flexible probes	1) CCCI of translations should include both source and target languages 2) Detected translation problems, cross-cutting problems, navigational problems, and English (source)-only problems
Goerman and Clifton (2011)	Determine useful-ness of vignettes for Hispanics within CCCI	English speakers Spanish bilinguals	1) English (66) 2) Spanish (57) (*n* = 123)	Items on housing from US Census	Retrospective probing Structured and flexible probes	1) Vignette approach was overall useful in both English and Spanish 2) Vignettes presented problems for Spanish speakers when they were not culturally appropriate

Table 1. *Continued*

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Hak, van der Veer, and Jansen (2008)	Test variant of cognitive testing: Three-Step Test Interview (TSTI), in two European countries	Norwegians Dutch	1) Norwegian 2) Dutch (*n* = unstated)	Items on alcohol consumption, attitudes toward illegal aliens, quality of life	Three-Step Test Interview (TSTI) procedure involving think-aloud and retrospective probes	TSTI cognitive interviewing procedure functioned effectively for Western/ Northern Europeans
Levin, Willis, Forsyth, Norberg, Stapleton, Stark, and Thompson (2009)	Evaluate CCCI in application to a Spanish-language translation	Spanish speakers English speakers –In 3 geographic areas of US	1) Spanish (27) 2) English (9) (*n* = 36)	Dietary (food frequency) questions	"Procedural pretest" followed by two iterative testing rounds Concurrent probes Flexible probes	1) CCCI in Spanish presented no significant obstacles 2) General/elaborative probe types were most effective: "Tell me more" and "Why did you answer like that?" 3) Cognitive testing detected translation, cultural, and generic problems

Continued

Table 1. *Continued*

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Mehrotra (2007)	Assess CCCI for use by English speakers in India	Residents of India	English ($n = 10$)	Emotional Processing Scale	Think-aloud; concurrent probing Flexible probes: paraphrasing, meaning probes	Cognitive interviewing concluded effective for English-speaking Indians, especially for identifying unclear survey questions
Miller (2003)	Assess comprehension of questions for poor with limited access to health care	Residents of rural Mississippi (21) Maryland (14) Ottawa (9)	English ($n = 44$)	Health survey questions	Concurrent probing Flexible probes	CI effective in illustrating problems involving question comprehension, background knowledge, and mental computation, for poor, survey-unacculturated subpopulation

Continued

Table 1. *Continued*

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Miller, Fitzgerald, Padilla, Willson, Widdop, Caspar, Dimov, Gray, Nunes, Prüfer, Schöbi, and Schoua-Glusberg (2011)	Develop and evaluate CCCI for multi-national testing for assessment of equivalence, identification of problematic items	1) US 2) UK 3) Bulgaria 4) Portugal 5) Switzerland 6) Germany 7) Spain	1) English, Spanish 2) English 3) Bulgarian 4) Portuguese 5) French 6) German 7) Spanish (*n* = 135; at least 10 per country)	Items on attitudes toward taxes, perceptions of age-status, and physical/mental health	Structured and flexible probes, Charting, collaborative analysis used	1) CCCI concluded effective 2) Collaborative analysis used to align results across countries 3) Use of structured matrix approach (chart) to portray interview-level results facilitates analysis
Miller, Mont, Maitland, Altman, and Madans (2011)	Apply CCCI in large multi-national context to assess equivalence and identify problematic items	15 countries across North America, South America, Asia, and Africa	Unspecified: Multiple (*n* = 1,290: 20–223 per country)	Disability, mainly involving vision	Concurrent probing Structured probes, with prior development	1) Structured probes are effective for novice interviewers, if carefully developed beforehand 2) Use of structured matrix (chart) approach to portray interview-level results facilitates analysis

Continued

Table 1. *Continued*

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Nápoles-Springer, Santoyo-Olsson, O'Brien, and Stewart (2006)	Evaluate CCCI and behavior coding in application to a patient survey	1) Hispanics 2) Non-Hispanic White, Black	1) Spanish and English (20) 2) English (28) (n = 48)	Patient survey	Retrospective probing Mixture of structured and flexible probes	1) No mention of problems with CCCI for Hispanics, or in Spanish 2) The combination of behavior coding and cognitive testing is effective in CCCI studies
Pan, Landreth, Park, Hinsdale-Shouse, and Schoua-Glusberg (2010)	Evaluation of CCCI in testing translations of survey materials	Monolinguals in: 1) English 2) Chinese 3) Korean 4) Russian 5) Spanish	1) English (16) 2) Chinese (24) 3) Korean (24) 4) Russian (25) 5) Spanish (24) (n = 113)	American Community Survey Advance Materials (letter and brochure)	Mix of structured and flexible probes Evaluated usefulness of responses to 7 probe types	Probe effectiveness (+/−) 1) Paraphrasing (−) 2) Chinese/Korean: Evaluative, sensitivity, hypothetical probes (−) 3) Russian/Spanish: Sensitivity, hypothetical probes (+) Koreans: Brief responses to probes

Continued

Table 1. *Continued*

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Park, Sha, and Pan (2013)	To evaluate CCCI within a test of a Korean questionnaire translation	Native English speakers, native Korean speakers; in two areas of US	1) English (16) 2) Korean (23) (*n* = 39)	2010 US Census questionnaire	Two testing rounds Structured and flexible probes	1) CCCI effective in Korean, producing similar frequencies and distributions of problems as in English 2) Translation problems were frequently identified in the Korean version
Pasick, Stewart, Bird, and D'Onofrio (2001)	Apply CCCI and focus groups to evaluate cancer-related items	1) African American 2) Chinese 3) Hispanics 4) Vietnamese	1) English (30) 2) Mandarin/ Cantonese (30) 3) Spanish (30) 4) Vietnamese (30) (*n* = 120)	Demographics, cancer-related practices and attitudes	Think-aloud Authors do not state whether probing was done	Thinking aloud was difficult for those with low educational level
Potaka and Cochrane (2004)	Use CCCI to test questionnaire on language use by Māori	Māori residents of New Zealand	English (24) Māori (18) (*n* = 42)	Items concerned the use of the Māori language	Type of probing procedures used is unstated	No indication of problems with CCCI for Māori

Continued

Table 1. *Continued*

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Reeve, Willis, Shariff-Marco, Breen, Williams, Gee, Alegria, Takeuchi, Stapleton, and Levin (2011)	Apply CCCI and psychometric methods to identify problems	1) Asian (6) 2) Hispanic (6) 3) African American (6) 4) American Indian (6) 5) Non-Hispanic White (6)	English ($n = 30$)	Questions on self-reported racial/ethnic discrimination	One testing round Concurrent probing Mix of structured and flexible probes	1) No difficulties with CCCI reported 2) Mixed-method approach involving Item Response Theory (IRT) from psychometrics and cognitive testing was useful in identifying and interpreting problems with survey questions
Ridolfo and Schoua-Glusberg (2011)	Assess equivalence of comprehension across English and Spanish	Primary language English Primary language Spanish	1) English (27) 2) Spanish (30) ($n = 57$)	Items concerning self-reported race and ethnicity	Retrospective probing Probing flexibility unstated	1) Hispanics, non-Hispanics processed items on race/ethnicity differently 2) No reference to problems related to use of probing within CCCI 3) Was potential demographic (educational level) confounding due to small scale of study

Continued

Table 1. *Continued*

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Sha and Pan (2013)	Evaluate methods for managing large CCCI study	1) Chinese speakers (129) 2) Korean speakers (139)	1) Mandarin (91) Cantonese (38) 2) Korean (139) ($n = 258$)	Census ACS Community Guide	Concurrent and retrospective probing Mix of structured, flexible probes	1) CCCI effective if experienced CIs are used 2) Problems were found with source questionnaire, so modification was required to resolve these
Thompson, Willis, Thompson, and Yaroch (2011)	Study of interpretive differences between cultural groups	Primary language: 1) English 2) Spanish 3) Korean 4) Chinese	1) English (54) 2) Spanish (55) 3) Korean (13) 4) Chinese (13) ($n = 135$)	Diet questions: Judgment of each fruit item as fruit versus vegetable	Retrospective probe requesting assignment of each item to fruit, vegetable, other, or DK	1) Identified significant differences in subgroup conceptualization concerning fruit versus vegetable 2) Simple judgment probe functioned well across groups
Thrasher, Quah, Dominick, Borland, Driezen, Awang, Omar, Hosking, Sirirassamee, and Boado (2011)	Use CCCI to identify problems with translations in multi-national study	1) US 2) Australia 3) Mexico 4) Uruguay 5) Thailand 6) Malaysia	1) English (20) 2) English (20) 3) Spanish (20) 4) Spanish (20) 5) Thai (20) 6) Malaysian (20) ($n = 120$)	Tobacco use items, psychosocial constructs	Structured probes	1) No problems with use of cognitive probes reported 2) Authors note that use of structured probing limits opportunities to identify problems in tested items

Continued

Table 1. *Continued*

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Warnecke, Ferrans, Johnson, Chapa-Resendez, O'Rourke, Chavez, Dudas, Smith, Schallmoser, Hand, and Lad (1996)	Evaluate a Quality of Life Index (QLI) for US African American and Mexican American cancer patients	US-based: African American, Mexican American cancer patients having high school education or less	English (23) Spanish (15) (*n* = 38)	Quality-of-Life Index for cancer patients	Concurrent probing Structured probes	1) No problems with CCCI reported 2) Low educational level was an impediment to question interpretation 3) Translation problems were identified within the Spanish version
Warnecke, Johnson, Chavez, Sudman, O'Rourke, Lacey, and Horm (1997)	Determine equivalence of cross-cultural cognitive processing of health survey questions	US-based: African American, Puerto Rican, Mexican American, non-Hispanic White	English (*n* = 423)	General health survey items	Structured probes used Hypothetical/projective probes used	1) No problems with CCCI reported 2) Concluded that African Americans and Hispanics were uncomfortable discussing sensitive behavior with interviewer from a different cultural group

Continued

Table 1. *Continued*

(a) Study	(b) Purpose	(c) Subgroups included	(d) Languages included	(e) Type of material studied	(f) Procedural features	(g) Major findings and conclusions
Willis (forthcoming)	Compare results of CCCI when same questionnaire tested	English speakers Spanish speakers Chinese speakers Korean speakers	1) English (67) 2) Spanish (45) 3) Mandarin/ 4) Cantonese (9) 5) Korean (27) ($n = 148$)	Perceptions of breast/ prostate cancer risk	Probes selected by each of four organiza- tions operating independently	1) No problems with CCCI reported by any of four organizations 2) The identical major questionnaire defect was detected across all languages
Willis and Zahnd (2007)	Assess cross-cultural equivalence between Koreans, non-Koreans	1) Monolingual Koreans 2) Bilingual Koreans 3) Non-Koreans	1) Korean (9) 2a) English (9) 2b) Korean (9) 3) English (9) ($n = 36$)	Health survey items	Think-aloud, concurrent probes Probe structuring unstated	1) No problems with CCCI reported 2) Differences in question function between groups were partly attributable to effects of acculturation to US society
Zeldenryk, Gordon, Gray, Speare, Melrose, Hossain, and Williams (2013)	Evaluate the Bangla- language version of a Quality of Life (QOL) scale	Residents of rural villages in Bangladesh having lymphatic filariasis	Bangla ($n = 35$)	The 26-item WHOQOL- BREF Quality of Life scale	A single testing round Think-aloud, concurrent probing Flexible probes	1) Evaluated QOL questions were too formal, irrelevant to participants, or poorly translated 2) Probing was effective, but think-aloud was not

NOTE.—Studies are published empirical investigations involving the conduct of CCCI, in sufficient detail to complete table elements (a)–(g).

Table 2. Example Findings from Cross-Cultural Cognitive Interviewing (CCCI) Studies

Study	Item tested	Description of findings from CCCI
1) Levin, Willis, Forsyth, Norberg, Stapleton, Stark, and Thompson (2009)	During the past month, how often did you eat…salsa?	*Problematic term*: In Spanish, "salsa" is a general term equivalent to "sauce" in English, so the question did not convey the intended meaning of a picante-like sauce. Spanish speakers reported thinking about marmalade, applesauce, and fruit sauce for topping ice cream.
2) Miller (2003)	Overall, in the last 30 days, how much difficulty did you have with work or household activities? Would you say none, mild, moderate, severe, or extreme?	*Problem with response category selection*: Low-income rural participants could indicate how much difficulty they had, in their own words, yet had difficulty selecting a one-word descriptor of the type provided (e.g., "…my housework… is hard…when I get ready to do something, I can't do it. If I get ready to dust, now, I can do it if I sit down on the floor, scooting around. I can do it that way").
3) Nápoles-Springer, Santoyo-Olsson, O'Brien, and Stewart (2006)	How often did doctors ask you if you wanted to include your family when making decisions?"	*Differential cultural conceptions of a concept between Hispanics and non-Hispanics*: For Spanish-speaking Latinos, family involvement was important to medical care; for non-Hispanic black/white respondents, the question was irrelevant, as they felt that involving family was often inappropriate.
4) Hak, van der Veek, and Jansen (2008)	How often did you drink six or more glasses on one day, during the last six months? [] Every day [] Never	*Problems in choosing response category due to variability in behavior*: Participants wanted to express variability of drinking behavior. Example: A German shift worker selected two response categories, and when probed explained that he only drinks alcohol in weeks in which he does not work. In such weeks he often drinks more than six glasses of beer a day, but this also varies.

Table 2. *Continued*

Study	Item tested	Description of findings from CCCI
5) Ridolfo and Schoua-Glusberg (2011)	How do other people usually see you in this country? Would you say people see you as Hispanic or Latino?	*Varying interpretation of item intent, between Hispanics and non-Hispanics:* Several Hispanics discussed whether they were discriminated against because they were Hispanic. Generally, some individuals also reported that the answer depends on whether the "other people" are themselves Hispanic or non-Hispanic, so this can be difficult to answer.
6) Thrasher, Quah, Dominick, Borland, Driezen, Awang, Omar, Hosking, Sirirassamee, and Boado (2011)	Tobacco is addictive (with 5-point agree/disagree Likert scale)	*Interpretations were varied across countries:* General control of behavior was a primary theme across all countries (Australia, United States, Uruguay, Mexico, Malaysia, and Thailand). However, frequencies of specific themes, including physiological and psychological dependency, frequency or quantity of consumption, and elements of pleasure versus danger, demonstrated differential cross-cultural patterns.

challenges that are somewhat unique to CCCI studies, both logistical and procedural. I will discuss these in turn.

LOGISTICAL ISSUES ASSOCIATED WITH CCCI

Establishment of sample size: One important development in the CCCI field, relative to standard cognitive testing studies, concerns the key design feature of participant sample size. In particular, due to the complexities related to the increased number of defined subgroups to be compared, CCCI studies often include large numbers of interviews, well beyond the general range of 15–30 cited in prior general reviews of cognitive interviewing (e.g., Willis 2005; Beatty and Willis 2007; Miller 2011). Of the 31 studies in table 1 that list total sample size, only six (19.4 percent) contained 30 or fewer participants, whereas 12 (38.7 percent) contained 31–100, and 13 (44.8 percent) included 101 or more. As a typical case, Goerman and Caspar (2010a) conducted 110 interviews to identify problems in the US Census form for various Hispanic groups. The ascendency of cross-cultural issues has evidently been associated with a trend toward a significant increase in quantitative scope, consistent with recommendations that sample size should be sufficient to reach saturation of results (Warnecke et al. 1996; Blair and Conrad 2011; Miller et al. 2014).

Selection of participants and interviewers: Beyond quantitative requirements imposed by increased sample size, CCCI studies—especially multilingual investigations—also pose qualitative logistical demands (Sha and Pan 2013). Significant attention has been paid to (a) the identification and enlistment of appropriate individuals to be interviewed; and (b) the selection and training of the cognitive interviewers.

(a) *Participant recruitment:* For CCCI studies, a universal challenge is the selection of participant recruitment criteria. For survey language translations, a fundamental determinant of recruitment strategy is whether interviews are to be conducted only of those speaking the target (translated) language(s), or of monolinguals in the source language as well (Goerman and Caspar 2010b). Researchers have sometimes evaluated only the target-language questionnaire, for example a Spanish translation but not the source English version (Agans, Deeb-Sossa, and Kalsbeek 2006; Goerman and Caspar 2010a: study 1). However, there appears to be an emerging consensus that source-language testing is vital—optimally in parallel with the target language(s), as opposed to sequentially (Potaka and Cochrane 2004; Tanzer 2005). The majority of studies in the literature that include translation assessment have chosen to include source-language cognitive interviews (Carter, Schoua-Glusberg, and Sha 2009; Goerman and Caspar 2010a: study 1, 2010b).

The most frequently made justification for source-language testing is that this provides a measure of baseline questionnaire functioning by which to assess the operation of the translation. As a related point, several authors have

independently noted that the testing of translated versions invariably suggests fundamental problems in the source; these problems have been variously referred to as *cross-cutting* (Goerman and Caspar 2010b), g*eneric* (Levin et al. 2009), or *problems in the source questionnaire* (Fitzgerald et al. 2011). Testing of source- as well as target-language versions is therefore necessary to establish whether these problems are truly general, as opposed to specific to the target version.

An associated issue concerning participant recruitment, for translated materials, is whether target versions should be administered to monolingual or bilingual speakers (Levin et al. 2009). There is no consensus on this issue, due to divergent perspectives concerning the relative merits of focusing testing efforts on single- versus dual-language use. Studies including recruitment of bilinguals (e.g., Willis and Zahnd 2007; Saleska, Kanya-Ngambi, and Alvarado 2009; Berrigan et al. 2010; Pan, Wake-Yelei, et al. 2014) have relied upon these as a bridge between monolingual speakers of the source and target languages. For example, based on the assumption that bilinguals are likely better acculturated to US society than are monolinguals, inclusion of bilingual Koreans allowed Willis and Zahnd (2007) to separate issues of language (English versus Korean) from those of acculturation level in influencing interpretation of health survey questions.

On the other hand, it is also common to advocate the recruitment of monolinguals (Levin et al. 2009; Park et al. 2013; Pan, Leeman, et al. 2014), based on the finding that such individuals tend to experience significant problems in completing survey questionnaires, and because they represent the group for whom the translated questionnaire must function (as bilinguals have the option of completion in English). This is a compelling argument, and suggests that recruitment of monolinguals is an advisable practice for translation testing. A complication is that the definition of monolingual status is not necessarily straightforward, and Park and Son (2014) have explored the effects of various screening criteria for selection of Chinese monolingual speakers. In some cases, participants selected as monolinguals may have some proficiency in the source language, which limits conclusions concerning functioning of the translation for pure monolinguals.

Recruitment to control demographic confounding: A final issue related to recruitment, for any study aiming to compare questionnaire functioning between subgroups (whether or not language translation is involved), is the degree to which investigators balance demographic characteristics between these groups. Most authors include a table summarizing the age, educational level, gender, and perhaps income level of participants within each defined subgroup. However, they often note that subgroups were not matched, such that variation in behavior across subgroups (either with respect to reactions to cognitive interviewing or to the tested materials) may be due partly to demographic (e.g., age) differences, as opposed to language or cultural group membership (Saleska, Kanya-Ngambi, and Alvarado 2009; Berrigan et al.

2010). Miller et al. (2005) did attempt to assess the independent effects of demographic factors by assessing frequency of problems identified between Hispanics and non-Hispanics through multiple regression analysis, and concluded that age, rather than Hispanicity, accounted for the dominant effects noted. The vast majority of qualitative CCCI studies, however, have lacked the opportunity to account for effects of demographic factors, and are therefore limited in their capacity for attributing findings uniquely to language or to subgroup membership. This observation is consistent with a call for greater use of mixed-methods approaches that include quantitative designs to supplement qualitative approaches, (Madans et al. 2011; Benítez and Padilla 2014), as these can facilitate the statistical assessment of a range of factors that influence question function.

Recruitment mechanisms for CCCI studies: Apart from the issue of *who* to recruit is that of *how* to recruit them, and CCCI studies often demand specialized approaches. Liu, Sha, and Park (2013) focused on recruitment sources for Asian participants, relying on several measures of success, including time efficiency (hours required per successful recruitment), outreach capacity (number of individuals reached), and eligibility rate (proportion of contacts that produce eligible participants). Liu, Sha, and Park (2013) report that newspaper advertisements were the most time efficient, and physical flyers the least; that outreach capacity was the highest for newspaper ads; and that eligibility rates were best for word-of-mouth recruitment. They conclude that no one source is optimal, and that CCCI studies should consider combinations of these.

(b) *Selection and training of cognitive interviewers:* A further critical challenge to CCCI studies is the establishment of an effective cognitive interviewing staff. This issue has been addressed mainly for translation testing, for which a clear requirement is facility in the target language(s), as well as the ability to communicate with members of the research team who are monolingual in the source language. In conjunction, these requirements typically demand bilingual language proficiency (e.g., Levin et al. 2009; Goerman and Caspar 2010a; Pan et al. 2010; Sha and Pan 2013). For multiple target languages, Goerman and Caspar (2010a) suggest that all interviewers should be fluent in the source language and in one additional target language.

   What CCCI studies rarely address, however, are interviewer characteristics other than language proficiency that may influence participant behavior in cognitive interviews. Especially for topics that are sensitive or private in nature, members of some cultures might be reticent to be interviewed by someone of another gender or cultural group. However, the opposite argument has also been made, suggesting that a "naïve outsider" will obtain the most useful information (Willis 2005). Within the CCCI domain, Goerman and Caspar (2010a) have concluded that a cultural outsider may be given additional latitude by the participant to ask probe questions that come across as

naïve, cumbersome, or inappropriate, or that otherwise violate conversational norms appropriate with members of one's own group. Such tendencies may vary with subgroup, however: Using a projective probing technique, Johnson et al. (1997), and Warnecke et al. (1997) reported that African Americans and Hispanics indicated greater discomfort in discussing sensitive topics with a different-culture interviewer than did white non-Hispanics. Overall, because relatively little attention has been paid to interviewer effects for any particular cultural group, this stands as an area ripe for attention.

Presumably, a major reason that CCCI investigators have not attended heavily to interviewer demographic characteristics is that they have been occupied with more pressing challenges in locating and training appropriate cognitive interviewers for translation testing. Of course, control over interviewer selection may be limited, as for multinational studies where staffing decisions are made exclusively by in-country collaborators. However, there appears to be widespread agreement that, as well as being fluent in the target language, interviewers optimally also have experience in translation, cognitive or qualitative interviewing, and survey research methods generally (Sha and Pan 2013). This needle-in-a-haystack requirement, along with associated costs, has led to several attempts to hire otherwise inexperienced bilinguals, and to compensate for survey inexperience by constraining the cognitive interviewing task so as to minimize its complexity and training requirements.

However, the practice of settling for otherwise inexperienced bilingual speakers as interviewers has sometimes proved insufficient. Pasick et al. (2001) reported problems with the conduct of non-English cognitive interviews, and this may be in part traced to their use of bilingual graduate students rather than seasoned professionals as interviewers. Further, Forsyth et al. (2007) relied on a highly experienced Survey Language Consultant (SLC) to hire and train two bilingual cognitive interviewers for each target language, yet concluded that it would be more effective to employ the SLCs as the cognitive interview staff. As a positive development, whereas a decade ago it was very difficult to locate cognitive interviewers who were bilingual, bicultural, and had prior experience in cognitive interviewing, a cadre of capable cognitive interviewers who are well versed in cognitive research appears to have more recently emerged for Spanish and several Asian languages. Several recent studies, such as those by Levin et al. (2009), Goerman and Clifton (2011), and Ridolfo and Schoua-Glusberg (2011), have included cognitive interviewers with prior training and experience in target-language cognitive testing, or who are described as study researchers integrated fully into the testing and analysis processes.

PROCEDURAL ISSUES IN CCCI

Cognitive interviewing studies feature a wide variety of procedural elements, but a key factor that may influence success is the type of cognitive interviewing technique applied. Following Beatty and Willis (2007), the most fundamental

divide is between think-aloud and verbal probing. Unfortunately, even the vetted, peer-reviewed studies within table 1 do not universally make clear the amount of think-aloud that was attempted or obtained. Further, use of terminology in the cognitive testing field is uneven; authors may use "think-aloud" as a general descriptor for a cognitive interview that also includes probing (e.g., Pasick et al. 2001), and it is unclear whether participant difficulties can be traced to failures to freely think aloud. Even where thinking aloud is specifically referenced, judgments concerning its use are varied. Levine et al. (2004) reported that think-aloud functioned well for Spanish speakers. On the other hand, Pan (2004) cited problems in its use with members of Asian cultures in particular, noting that there is no direct translation of "thinking out loud" in Chinese. Further, within a study involving US African Americans, Hispanics, Chinese, and Vietnamese, Pasick et al. (2001) reported that thinking aloud presented challenges for participants with low educational levels, and Zeldenryk et al. (2013) obtained a similar result in rural Bangladesh.

More systematic attention to the establishment of whether thinking aloud presents particular problems for particular subgroups would be helpful. However, based on current findings suggesting that group-specific difficulties with think-aloud procedures could produce confounding when comparing testing results, it may be best to advise CCCI researchers to not rely solely on think-aloud, and to be suspicious of between-group differences in apparent target-question function that derive from its use. Rather—especially given oft-cited difficulties that many participants have with thinking aloud generally (Willis 2005)—it may be more appropriate to focus on verbal probing, as has been a trend in the standard cognitive testing literature.

Even if one accepts that targeted verbal probing by the interviewer is appropriate for CCCI studies, there has been debate concerning the optimal nature of these probes: Should they be standardized and scripted in order to decrease interviewer variance and to facilitate interviewing by novice interviewers, or flexible and unscripted to take advantage of the inherent adaptability of cognitive interviewing? In brief, based on evolutionary developments of the past 15 years, there appears to be considerable agreement in the CCCI field that flexible rather than structured probing is desirable—even among some authors who have chosen a more standardized approach (e.g., Thrasher et al. 2011). Overall, of the 23 studies within table 1 for which probing strategy was clear, 17 (73.9 percent) involved at least some flexible probes. It may be feasible to rely on structured probes, if the investigation is so large that there is an incentive to align the probing for ease of analysis (Miller et al. 2011). However, trade-offs still apply: For effective use by inexperienced interviewers, probes must be very carefully developed, and even pretested themselves (Levin et al. 2009; Miller et al. 2011). What seems to work less well is to treat bilingual interviewers as, in effect, automatons that administer directly translated, standardized probes in word-for-word fashion.

Types of probes that function in particular subgroups: Even under flexible administration, it is possible that cognitive probing differs in its efficacy across linguistic or cultural groups (table 3 contains a compendium of probe types commonly used in both standard cognitive testing and CCCI studies). In parallel to arguments made above concerning thinking aloud, the possibility that probes may be problematic in application to particular subgroups presents a major potential threat to comparative CCCI investigations. Several reports have referred to problems with the use of specific cognitive probe types by both Hispanics and Asians in the United States (e.g., Kissam, Herrera, and Nakamoto 1993; Carrasco 2003; Pan 2004; Yuan et al. 2009). However, based on the total set of published studies abstracted for this review (table 1), the accumulated evidence does not support the notion that there is any culture for which cognitive probing is ineffective, and the vast majority of studies have concluded that probing works well across the full linguistic and cultural spectrum.

There is evidence that common probe types do vary in their effectiveness within CCCI studies (Goerman and King 2014). Problems associated with the use of paraphrasing have been identified (Pan 2004; Pan, Wake-Yelei, et al. 2014), but this again reiterates the general finding that paraphrase probes tend to be difficult (Prüfer and Rexroth 2003). Further, probes that ask for opinions

Table 3. Categories of Cognitive Probes used in Cross-Cultural Cognitive Interviewing (CCCI), Based on Pan, Wake-Yelei, Chan, and Willis (2014) and Willis (2005)

1) Comprehension/Interpretation/Meaning-oriented probe
"What, to you, is 'ethnic origin'?"
2) Paraphrase
"Can you tell me in your own words what this question is asking?"
3) Process-oriented probe
"How did you arrive at that answer?"
4) Confidence judgment
"*How sure are you that your total household income was less than $40,000?"*
5) Evaluative probe
 "Do you feel this question is easy or difficult to answer?"
6) Elaborative probes
 "Why do you say that?"; "Tell me more"
7) Hypothetical probe
 "How would you answer this question if your son lived at home less than half of the time?"
8) Recall probe
 "How do you remember the last time you visited a health professional?"
9) Sensitivity probe
"Do you think that this question asks about things that are too private, or is it ok to ask this?"

about survey questions or other materials violate the fundamental premise that we do not regard our cognitive testing participants as experts, and that it is ultimately up to the researcher to identify flaws (Willis 2005).

Concerning a finer-level analysis of probe function, Pan (2014) reported that in contrast to paraphrasing, process-oriented probes ("How did you come up with that answer?") and meaning-oriented probes ("What does the term/ phrase X mean to you in this question?") worked well in Chinese. There is one published study (Pan et al. 2010) that addressed probe function systematically for both Asian and Spanish speakers, and reported that evaluative, sensitivity, and hypothetical probes were relatively ineffective for Chinese and Koreans. Goerman and Clifton (2011) and Sha and Pan (2009) found that vignettes describing detailed hypothetical scenarios are effective with non-English participants, as long as the situation described by the vignette is not overly complex and does not violate cultural conventions. Finally, there has been less focus on general, elaborative probes (Willis 2005) such as "Tell me more about that," which are used to produce what Miller et al. (2014) refer to as the *narrative* comprising the fundamental basis for cognitive interviewing analysis, although Levin et al. (2009) reported that elaborative probes were especially effective with Spanish speakers.

A related issue, apart from the reactions to particular probe types, pertains to culturally associated normative response styles (Park, Sha, and Pan 2013). Chan (2010) found that Chinese participants were more likely than English speakers to provide brief and contradictory, Contrary-to-Face-Value (CTFV) responses (41 versus 0 percent of participants, respectively), in response to the hypothetical probe "If you were selected, would you participate in the survey?" Surely such tendencies can influence behavior within cognitive interviews, and may call into question the results of studies where responses to probes are insufficient. Again, however, this phenomenon is not unique to any particular culture, as it is commonly found that probes may be misunderstood (Blair and Piccinino 2005). In such cases, the interviewer is encouraged to make use of non-standardized probing techniques to follow up flexibly, and to rephrase or substitute probes in order to obtain the information desired (Wellens 1994). This requirement again speaks to the importance of involving cognitive interviewers who are experienced and knowledgeable concerning the measurement objectives of the tested items, so that they can interview effectively even where this requires unscripted follow-up probing in particular cultures or languages (Zeldenryk et al. 2013).

Other facets of cognitive probing: There is some debate in the standard cognitive testing arena concerning the usefulness of concurrent probes (those administered immediately after administration of each tested item) versus retrospective probing (debriefing following administration of all tested items) (Willis 2005). Both varieties are represented within the studies depicted in table 1: Of the 22 studies for which this was clear, 11 included concurrent probes; nine retrospective,

and two both. However, none of these mentioned the relative merits of either approach in application to CCCI. Further, concerning the general communication of the cognitive testing task to participants, several authors have suggested that recent immigrants may lack sufficient survey literacy (Agans, Deeb-Sossa, and Kalsbeek 2006; Chan and Pan 2011); if they do not understand the purpose of a survey, they are unlikely to be effective cognitive testing participants. However, the same argument has been made concerning low-income US citizens having little survey experience (Miller 2003). In all these cases, it is vitally important that the cognitive testing participant understand the ultimate purpose both of a survey and of the cognitive interview itself (Chan and Pan 2011). What have not been developed are optimal procedures or scripts that function to convey these messages to the survey unacculturated.

ANALYTIC ISSUES IN CI

The most undeveloped area of cognitive interviewing methodology has been analysis of the results (Miller 2011; Willis forthcoming). This observation pertains to the CCCI area as well, and is exacerbated by the inclusion of the explicit analysis levels necessitated when multiple linguistic or cultural groups are included. To delineate current practices, I will in turn review (a) data-reduction procedures; (b) language translation of testing results; (c) application of coding schemes; and (d) use of data displays as aids to analysis.

Data-reduction procedures: Data from cognitive interviews consist of either verbatim transcripts (Nápoles-Springer et al. 2006) or written interviewer notes that are made either during or after the interview. Reducing these to the level of summary information necessary to reach overall interpretations and conclusions has generally depended on two major approaches. The first, referred to as successive aggregation (Willis 2015), makes use of several hierarchical stages of summarization that are often accomplished by different individuals at each stage. For example, a series of Spanish-language interviews might be summarized first by each individual interviewer, and then those results further aggregated (and perhaps translated into English at this point) by a Spanish-speaking team member to represent the Spanish-language interviews. These text summaries could then be contrasted with results based on compilations of English-language interviews, by a monolingual lead investigator.

The successive aggregation of interview results has been advocated for multilingual investigations (Lee 2014), and has resulted in some positive benefits. A study of a self-administered measure of perceptions of cancer risk, described by Willis (2015), involved parallel, independent cognitive testing and analysis across four cognitive labs, and once aggregated and compared by the lead researcher, the findings were virtually identical across English, Spanish, Chinese, and Vietnamese groups, leading to the conclusion that the instrument contained a fundamental formatting flaw that resulted in gross

misinterpretation. In this case, the separate, successively aggregated analyses pointed to a coherent, mutually reinforcing conclusion. A potential pitfall to this approach, however, is that different analysts are reviewing different results, such that discrepancies between analysis approaches may be confounded with language or cultural group, making direct comparisons difficult. Miller (in Willis and Miller 2008) describes a failure of this technique, where cognitive testing results varied markedly from different countries, pointed to inconsistent and conflicting conclusions, and rendered the results uninterpretable.

As a solution, Miller et al. (2011) advocated a collaborative (*Joint*) analysis approach that depicts every result pertaining to each tested question at the lowest (interview) level, forcing all analysts to review the results prior to any further processing, interpretation, or aggregation. The advantage to joint analysis is that it avoids the possibility of analyst-dependent bias due to separate, uncoordinated data reduction and interpretation. On the other hand, the process of joint review of individual interviewing results by all key study personnel can be burdensome: Miller et al. (2011) report that even after a full three-day analysis meeting, the results of one study required significant further processing by the lead researcher. A compromise view is that, no matter what variety of data reduction is selected, separate interviewing teams should ideally not work independently, but rather feature a high degree of ongoing communication and sharing of information concerning results at multiple points (e.g., Reeve et al. 2011).

Translation of testing results: As a subsidiary issue in analysis of results of cognitive testing of translations, it is not clear at what point it is best to convert cognitive-testing results from target-language interviews back into the source language for consumption by (monolingual) project leaders. The approaches that have been used range from conducting the interview in the target language and simultaneously taking notes in the source language, to writing all interview summaries in the target language and translating only a summary into the source (Chan and Pan 2011). Goerman and Caspar (2010a) describe a variant in which interview reports are written in the source language, but with critical examples and language-critical results also listed in the target language. Overall, it seems more critical to retain original language expression when issues of lexical translation dominate, as opposed to conceptual issues that are language independent.

Coding of results in CCCI studies: A further issue related to analysis of cognitive interviews is the coding of the results. Standard cognitive interviewing had tended to eschew the application of codes and instead makes use of text-based descriptions of the findings (Willis forthcoming). In the CCCI domain, studies have increasingly made use of coding schemes, although these systems differ significantly. For studies categorizing the types of problems that exist within translated instruments, there has been a convergence in coding approaches, even by researchers working independently. Willis and Zahnd (2007) determined

that the results of cognitive interviews conducted in Korean and English could generally be classified as (1) translation problems; (2) problems of cultural adaptation; or (3) generic problems of questionnaire design. An almost identical system was introduced by Fitzgerald et al. (2011), who further sub-divided translation problems to distinguish translation error from translation difficulty. Other coding schemes for use in CCCI are either slightly more extensive (Pan and Fond 2011) or much more elaborate (e.g., Lee 2014).

The distinction between problems due to language translation, as opposed to those related to socio-cultural differences, seems especially prominent, and supports the assertion that CCCI studies should target not only translation issues, but any form of cultural variation that may influence survey response. A pointed example is that by Priede et al. (2010), who found that usage of a 0–10 scale by Finnish participants differed fundamentally from that by residents of other countries, and this was attributed not to linguistic issues, but rather to Finns' interpretation of a "4" response as "failure," based on experience with the school grading system unique to Finland. Such effects may be especially pronounced within cross-national CCCI studies that present social- and structural-system variation, in addition to the myriad cultural factors that exist within studies limited to a single nation.

In departure from an emphasis on problem characterization via application of an a priori coding scheme, some studies that assess cross-cultural equivalence tend to rely on a grounded theory approach that emphasizes the construction of codes from the available data (Daveson et al. 2011; Ridolfo and Schoua-Glusberg 2011; Thrasher et al. 2011). For example, Behr et al. (2014) coded open-ended responses to a probe embedded in a field survey ("What ideas do you associate with the phrase 'civil disobedience'? Please give examples"). The advantage to such inductive, "bottom-up" codes is that they are driven by the data to provide an unbiased assessment of divergence of interpretation across groups. On the other hand, these customized codes are by their nature specific to the item evaluated, and are not transportable to other items, contexts, or studies, as are predefined coding categories.

Use of data displays as analysis aids: The use of data displays including charts, matrices, or templates has been advocated as an alternative to the more unguided, open-ended write-up of summary notes, especially for larger studies (Miller et al. 2011). Further, a column-oriented representation of the results can be used as a guide to data collection initially (as through structured probing), if each column heading specifies a critical item of information to be collected during the interview. Miller et al. (2011) describe the use of data displays to identify response patterns that have diagnostic value in assessing question function: Probes that accompanied items on vision problems were used to ascertain inconsistent response patterns that revealed, for example, that many participants failed to encode a critical part of an item instructing them to include the wearing of glasses when self-assessing their visual acuity.

Conclusions: Potentially Effective Practices for CCCI

To summarize the review, I reiterate the major points suggested by the set of CCCI reports reviewed. Given the paucity of studies that actively investigate each of these issues, these are presented as working hypotheses, as opposed to data-driven conclusions:

(1) In assessing cross-cultural equivalence, researchers should consider recruiting significantly more participants than for a simple, standard cognitive interviewing study. In doing so, it is helpful to treat participant recruitment as a special challenge involving the hard-to-reach, making use of varied and targeted approaches, including outreach to community groups, internet-based social media, or other avenues that may be particularly well suited to the culturally or linguistically isolated.

(2) With respect to interviewer selection and training, there is, as the saying goes, no free lunch: For testing of translations in particular, it may be advisable to either make use of sophisticated interviewers to enable probing flexibility, or else to hire less experienced speakers of the target language who are capable of administering standard probes, but attending extensively to probe development.

(3) Investigators should anticipate that CCCI of translations will reveal potential problems with the source-language questionnaire. As a consequence, additional testing of the source version may be necessary to confirm these findings.

(4) Cognitive probing appears to be effective for all cultural and language groups studied to date. Probe varieties that have been found to be problematic (e.g., paraphrasing) are likely to present difficulties for immigrants, the unacculturated, those with lower levels of education, and members of cultures with communication styles that depart from that in which cognitive interviewing was developed.

(5) To the degree possible, probing should be flexible rather than completely standardized. If structured probes are used, it is helpful to be able to follow up with more flexible, spontaneous probing.

(6) Analysis is facilitated when observed differences in cognitive testing results between subgroups can be attributed to group membership, rather than to the nature of the analysis method. A promising approach to achieving analysis comparability is a joint, collaborative procedure in which all researchers are involved in making determinations concerning question functioning at the individual interview level.

Recommendations for Future Research and Practice Related to CCCI

Overall, CCCI appears to be an effective process within cross-cultural and multilingual applications. However, the evidence for this statement could be

strengthened through several practices. One major recommendation concerns reporting findings: Even some of the studies in table 1, which were selected on the basis of containing sufficient procedural information, still contained significant gaps with respect to key variables such as whether iterative testing was done, type of probing approach (use of think-aloud, concurrent versus retrospective, degree of scripting of probes), prior experience of cognitive interviewers, type and length of training, decision rules concerning whether saturation of results has been achieved, and nature of compilation of testing results. To increase transparency and facilitate future reviews, studies should clarify each of these elements, by making use of a checklist-based reporting framework such as the Cognitive Interviewing Reporting Format (CIRF) introduced by Boeije and Willis (2013). Finally, I call upon researchers to not only describe applications of CCCI, but to build into their investigations explicit evaluation elements and metrics. In particular, investigations could focus on: (a) assessment of recruitment procedures for monolinguals and hard-to-reach subgroups; (b) effects of interviewer characteristics and behavior; (c) effectiveness of alternate probe models and types; (d) selection of data reduction and analysis procedures, for cross-national as well as general cross-cultural contexts; and (e) indicators of process quality, such as measurement of inter-rater reliability of code assignment. By attending to these methodological details, CCCI methods can be better critiqued as we move toward best practices in this endeavor.

References

Agans, R. P., Deeb-Sossa, N., and William D. Kalsbeek. 2006. "Mexican Immigrants and the Use of Cognitive Assessment Techniques in Questionnaire Development." *Hispanic Journal of Behavioral Sciences* 28:209–30.

Beatty, Paul, and Gordon Willis. 2007. "Research Synthesis: The Practice of Cognitive Interviewing." *Public Opinion Quarterly* 71:287–311.

Behr, Dorothee, Michael Braun, Lars Kaczmirek, and Wolfgang Bandilla. 2014. "Item Comparability in Cross-National Surveys: Results from Asking Probing Questions in Cross-National Surveys About Attitudes Towards Civil Disobedience." *Quality and Quantity* 48:127–48.

Benítez, Isabel., and Jose-Luis Padilla. 2014. "Analysis of Nonequivalent 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* 8:52–68.

Berrigan, David, Barbara Forsyth, Cynthia Helba, Kerry Levin, Alicia Norberg, and Gordon Willis. 2010. "Cognitive Testing of Physical Activity and Acculturation Questions in Recent and Long-Term Latino Immigrants." *BMC Public Health* 10. Available at http://www.biomedcentral.com/1471–2458/10/481.

Blair, Johnnie, and Frederick G. Conrad. 2011. "Sample Size for Cognitive Interview Pretesting." *Public Opinion Quarterly* 75:636–58.

Blair, Johnny, and Linda Piccinino. 2005. "The Development and Testing of Instruments for Cross-Cultural and Multi-Cultural Surveys." In *Methodological Aspects in Cross-National Research,* edited by Jurgen H. P. Homeyer-Zlotnik and Janet Harkness, 13–30. Mannheim, Germany: ZUMA Nachrichten Spezial.

Boeije, Hennie, and Gordon Willis. 2013. "The Cognitive Interviewing Reporting Framework (CIRF): Towards the Harmonization of Cognitive Interviewing Reports." *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 9:87–95. doi: 10.1027/1614–2241/a000075.

Carrasco, Lorraine. 2003. "The American Community Survey (ACS) en Español: Using Cognitive Interviews to Test the Functional Equivalency of Questionnaire Translations." *Statistical Research Division Research Report Series* (Survey Methodology #2003–17). Washington, DC: US Census Bureau.

Carter, George R., Alisu Schoua-Glusberg, and M. Mandy Sha. 2009. "Language, Culture, and Respondent Knowledge: Findings from the Cognitive Test of the Spanish Translation of the American Housing Survey." *Proceedings of the Survey Methods Research Section of the American Statistical Association*, 5925–5939.

Chan, Anna Y. 2010. "Analysis of Chinese Speakers' Responses to Survey Intention Questions." *Proceedings of the Survey Methods Research Section of the American Statistical Association*, 1672–1686.

Chan, Anna Y., and Yuling Pan. 2011. "The Use of Cognitive Interviewing to Explore the Effectiveness of Advance Supplemental Materials Among Five Language Groups." *Field Methods* 23:342–61.

Daveson, Barbara A., Dorothee Bechinger-English, Claudia Bausewein, Steffan T. Simon, Richard Harding, Irene J. Higginson, and Barbara Gomes. 2011. "Constructing Understandings of End-of-Life Care in Europe: A Qualitative Study Involving Cognitive Interviewing with Implications for Cross-National Surveys." *Journal of Palliative Medicine* 14:343–49.

Fitzgerald, Rory, Sally Widdop, Michele Gray, and Deborah Collins. 2011. "Identifying Sources of Error in Cross-National Questionnaires: Application of an Error Source Typology to Cognitive Interview Data." *Journal of Official Statistics* 27:569–99.

Forsyth, Barbara H., Martha S. Kudela, Kerry Levin, Deirdre Lawrence, and Gordon B. Willis. 2007. "Methods for Translating an English-Language Survey Questionnaire on Tobacco Use into Mandarin, Cantonese, Korean, and Vietnamese." *Field Methods* 19:264–83.

Fujishiro, Kaori, Fang Gong, Sherry Baron, C. Jeffery Jacobson Jr., Sheli DeLaney, Michael Flynn, and Donald E. Eggerth. 2010. "Translating Questionnaire Items for a Multi-Lingual Worker Population: The Iterative Process of Translation and Cognitive Interviews with English-, Spanish-, and Chinese-Speaking Workers." *American Journal of Industrial Medicine* 53:94–203.

Gerber, Eleanor R. 1999. "The View from Anthropology: Ethnography and the Cognitive Interview." In *Cognition and Survey Research*, edited by Monroe Sirken, Douglas Herrmann, Susan Schechter, Norbert Schwarz, Judy Tanur, and Roger Tourangeau, 217–34. New York: Wiley.

Goerman, Patricia, and Rachel Casper. 2010a. "Managing the Cognitive Pretesting of Multilingual Survey Instruments: A Case Study of Pretesting of the US Census Bureau Bilingual Spanish/English Questionnaire." In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Janet Harkness, Michael Braun, Brad Edwards, Tim Johnson, Lars Lyberg, Peter Mohler, Beth E. Pennell, and Tom Smith, 75–90. New York: Wiley.

———. 2010b. "A Preferred Approach for the Cognitive Testing of Translated Materials: Testing the Source Version as a Basis for Comparison." *International Journal of Social Research Methodology* 13:303–16.

Goerman, Patricia L., and Matthew Clifton. 2011. "The Use of Vignettes in Cross-Cultural Cognitive Testing of Survey Instruments." *Field Methods* 23:362–78.

Goerman, Patricia, and Ryan King. 2014. "Adaptation of Standard Cognitive Interview Methodology for Use with Spanish-Speaking Respondents." Paper presented at the 69th Annual Conference of the American Association for Public Opinion Research (AAPOR), Anaheim, CA, USA.

Hak, Tony, Kees van der Veer, and Harrie A. M. Jansen. 2008. "The Three-Step Test-Interview (TSTI): An Observation-Based Method for Pretesting Self-Completion Questionnaires." *Survey Research Methods* 2:143–50.

Harkness, Janet A., Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom Smith. 2010. *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Hoboken, NJ: Wiley.

Harkness, Janet A., Brad Edwards, Sue Ellen Hansen, Deborah R. Miller, and Anna Villar. 2010. "Designing Questionnaires for Multipopulation Research." In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Janet A. Harness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom Smith, 33–57. Hoboken, NJ: Wiley.

Johnson, Timothy P. 2006. "Methods and Frameworks for Crosscultural Measurement." *Medical Care* 44:S17–20.

Johnson, Timothy, Diane O'Rourke, Noel Chavez, Seymour Sudman, Richard Warnecke, Loretta Lacey, and John Horm. 1997. "Social Cognition and Responses to Survey Questions among Culturally Diverse Populations." In *Survey Measurement and Process Quality*, edited by Lars Lyberg, Paul Biemer, Martin Collins, Edith De Leeuw, Catherine Dippo, Norbert Schwarz, and Dennis Trewin, 87–113. Hoboken, NJ: Wiley.

Kissam, Edward, Enrique Herrera, and Jorge M. Nakamoto. 1993 (March). "Hispanic Response to Census Enumeration Forms and Procedures." Report submitted to the US Census Bureau, Center for Survey Methods Research. Contract No. 50-YABC-2-66027, Task Order No. 46-YABC-2-0001.

Lee, Jihyun. 2014. "Conducting Cognitive Interviews in Cross-National Settings." *Assessment* 21:227–40.

Levin, Kerry, Gordon W. Willis, Barbara H. Forsyth, Alicia Norberg, Martha K. Stapleton, Deborah Stark, and Frances E. Thompson. 2009. "Using Cognitive Interviews to Evaluate the Spanish-Language Translation of a Dietary Questionnaire." *Survey Research Methods* 3:13–25.

Levine, Roger, Raquel González, Beverly Weidmer, and Patricia Gallagher. 2004. "Cognitive Testing of English and Spanish Versions of Health Survey Items." *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 4818–4825.

Liu, Lu, M. Mandy Sha, and Hyunjoo Park. 2013. "Exploring the Efficiency and Utility of Methods to Recruit Non-English Speaking Qualitative Research Participants." *Survey Practice [S.l.]*6, ISSN 2168-0094. Available at http://www.surveypractice.org/index.php/SurveyPractice/article/view/56.

Madans, Jennifer, Kristen Miller, Aaron Maitland, and Gordon Willis. 2011. *Question Evaluation Methodology*. Hoboken, NJ: Wiley.

Mehrotra, Seema. 2007. "Cognitive Interviewing: An Overview and an Illustration." *Journal of the Indian Academy of Applied Psychology* 33:81–84.

Miller, Kristen. 2003. "Conducting Cognitive Interviews to Understand Question-Response Limitations among Poorer and Less-Educated Respondents." *American Journal of Health Behavior* 27(S3):264–72.

———. 2011. "Cognitive Interviewing." In *Question Evaluation Methods*, edited by Jennifer Madans, Kristen Miller, Aaron Maitland, and Gordon Willis, 51–75. Hoboken, NJ: Wiley.

Miller, Kristen, Rory Fitzgerald, José-Luis Padilla, Stephanie Willson, Sally Widdop, Rachel Caspar, Martin Dimov, Michele Gray, Cátia Nunes, Peter Prüfer, Nicole Schöbi, and Alisú Schoua-Glusberg. 2011. "Design and Analysis of Cognitive Interviews for Comparative Multinational Testing." *Field Methods* 23:379–96.

Miller, Kristen, Daniel Mont, Aaron Maitland, Barbara Altman, and Jennifer Madans. 2011. "Results of a Cross-National Structured Cognitive Interviewing Protocol to Test Measures of Disability." *Quality and Quantity* 45:801–15.

Miller, Kristen, Gordon Willis, Connie Eason, Lisa Moses, and Beth Canfield. 2005. "Interpreting the Results of Cross-Cultural Cognitive Interviews." In *Methodological Aspects in Cross-National Research*, edited by Jurgen H. P. Homeyer-Zlotnik and Janet Harkness, 79–92. Mannheim, Germany: ZUMA Nachrichten Spezial.

Miller, Kristen, Stephanie Willson, Valerie Chepp, and Jose-Luis Padilla. 2014. *Cognitive Interviewing Methodology*. New York: Wiley.

Nápoles-Springer, Anna M., Jasmine Santoyo-Olsson, Helen O'Brien, and Anita L. Stewart. 2006. "Using Cognitive Interviews to Develop Surveys in Diverse Populations." *Medical Care* 44(S3):S21–30.

Pan, Yuling. 2004. "Cognitive Interviews in Languages Other Than English: Methodological and Research Issues." *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 4859–4865.

Pan, Yuling, and Marissa Fond. 2011. "Evaluating Multilingual Questionnaires: A Sociolinguistic Perspective." *Research and Methodology Directorate, Center for Survey Measurement Study Series* (Survey Methodology #2012-04). US Census Bureau. Available at http://www.census.gov/srd/papers/pdf/ssm2012-04.pdf.

Pan, Yuling, Ashley Landreth, Hyunjoo Park, Marjorie Hinsdale-Shouse, and Alisu Schoua-Glusberg. 2010. "Cognitive Interviewing in Non-English Languages: A Cross-Cultural Perspective." In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg, Peter Mohler, Beth E. Pennell, and Tom W. Smith, 91–113. Hoboken, NJ: Wiley.

Pan, Yuling, Jennifer Leeman, Marissa Fond, and Patricia Goerman. 2014. "Multilingual Survey Design and Fielding: Research Perspectives from the US Census Bureau." *Research Report Series* (Survey Methodology #2014-01).

Pan, Yuling, Virginia Wake-Yelei, Grace Chan, and Gordon B. Willis. 2014. "A Comparative Study of English and Chinese Cognitive Interviews." Paper presented at the Comparative Survey Design and Implementation Workshop, Bethesda, MD, USA.

Park, Hyunjoo, M. Mandy Sha, and Yuling Pan. 2013. "Investigating Validity and Effectiveness of Cognitive Interviewing as a Pretesting Method for Non-English Questionnaires: Findings from Korean Cognitive Interviews." *International Journal of Social Research Methodology.* Available at http://dx.doi.org/10.1080/13645579.2013.823002.

Park, Hyunjoo, and Jiyoung Son. 2014. "Identifying Monolingual Respondents for Cognitive Interviewing." *Field Methods* 26:269–83.

Pasick, Rena J., Susan L. Stewart, Joyce A. Bird, and Carol N. D'Onofrio. 2001. "Quality of Data in Multiethnic Health Surveys." *Public Health Reports* 116:223–43.

Potaka, Lynn, and Suzanne Cochrane. 2004. "Developing Bilingual Questionnaires: Experiences from New Zealand in the Development of the 2001 Maori Language Survey." *Journal of Official Statistics* 20:289–300.

Priede, Camilla, Elina Ruuskanen, Anniina Jokinen, and Stephen Farrall. 2010. "Analysing Cognitive Interview Data to Improve Cross-National Survey Questions." University of Surrey, Social Research Update, 59.

Prüfer, Peter, and Rexroth, Margrit. 2003. "Paraphrasing Can Be Dangerous: A Little Experiment." Paper presented at the Proceedings of the 4th Conference on Questionnaire Evaluation Standards (QUEST), Mannheim, Germany.

Reeve, Bryce B., Gordon Willis, Salma N. Shariff-Marco, Nancy Breen, David R. Williams, Gilbert C. Gee, Margarita Alegría, David T. Takeuchi, Martha Stapleton, and Kerry Levin. 2011. "Comparing Cognitive Interviewing and Psychometric Methods to Evaluate a Racial/Ethnic Discrimination Scale." *Field Methods* 23:397–419.

Ridolfo, Heather, and Alisu Schoua-Glusberg. 2011. "Analyzing Cognitive Interview Data Using the Constant Comparative Method of Analysis to Understand Cross-Cultural Patterns in Survey Data." *Field Methods* 23:420–38.

Saleska, Erica, Musindu Kanya-Ngambi, and Herman A. Alvarado. 2009. "How Does the French Culture Impact the Translation of Survey Materials? An Examination of French Translations for a US Survey." *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 5529–43.

Sha, M. Mandy, and Yuling Pan. 2009. "The Use of Vignettes in Evaluating Multilingual Questionnaires." *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 6122–6133.

———. 2013. "Adapting and Improving Methods to Manage Cognitive Pretesting of Multilingual Survey Instruments." *Survey Practice*, 6.

Smith, Tom W. 2004. "Developing and Evaluating Cross-National Survey Instruments." In *Methods for Testing and Evaluating Survey Questionnaires*, edited by Stanley Presser, Jennifer M. Rothgeb, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer, 431–52. Hoboken, NJ: Wiley.

Tanzer, Norbert K. 2005. "Developing Tests for Use in Multiple Languages and Cultures: A Plea for Simultaneous Development." In *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*, edited by Ronald K. Hambleton, Peter F. Merenda, and Charles D. Spielberger, 235–63. Mahwah, NJ: Erlbaum.

Thompson, Frances, Gordon B. Willis, Olivia M. Thompson, and Amy L. Yaroch. 2011. "The Meaning of 'Fruits' and 'Vegetables.'" *Public Health Nutrition* 14:1222–1228.

Thrasher, James F., Anne C. K. Quah, Gregory Dominick, Ron Borland, Pete Driezen, Rahmat Awang, Maizurah Omar, Warwick Hosking, Buppha Sirirassamee, and Marcelo Boado. 2011. "Using Cognitive Interviewing and Behavioral Coding to Determine Measurement Equivalence Across Linguistic and Cultural Groups: An Example from the International Tobacco Control Policy Evaluation Project." *Field Methods* 23:439–60.

Warnecke, Richard, Carol E. Ferrans, Timothy P. Johnson, Gloria Chapa-Resendez, Diane P. O'Rourke, Noel Chavez, Susan Dudas, Eva D. Smith, Lucy M. Schallmoser, Roger P. Hand, and Thomas Lad. 1996. "Measuring Quality of Life in Culturally Diverse Populations." *Journal of the National Cancer Institute Monographs* 20:29–38.

Warnecke, Richard B., Timothy P. Johnson, Noel Chavez, Seymour Sudman, Diane P. O'Rourke, Loretta Lacey, and John Horm. 1997. "Improving Question Wording in Surveys of Culturally Diverse Populations." *Annals of Epidemiology* 7:334–42.

Wellens, Tracey. 1994. "The Cognitive Evaluation of the Nativity Questions for the Current Population Survey." *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 1204–1209.

Willis, Gordon B. 2005. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage.

———. 2015. "Pretesting of Health Survey Questionnaires: Cognitive Interviewing, Usability Testing, and Behavior Coding." In *Handbook of Health Survey Methods*, edited by T. Johnson, 217–42. Thousand Oaks, CA: Sage.

———. Forthcoming. *Analysis of the Cognitive Interview in Questionnaire Design*. Cambridge: Oxford.

Willis, Gordon B., and Kristen Miller. 2008. "Analyzing Cognitive Interviews." Short course presented at the meeting of the Southern Association for Public Opinion Research, Durham, NC, USA.

———. 2011. "Cross-Cultural Cognitive Interviewing: Seeking Comparability and Enhancing Understanding." *Field Methods* 23:331–41.

Willis, Gordon, and Elaine Zahnd. 2007. "Questionnaire Design from a Cross-Cultural Perspective: An Empirical Investigation of Koreans and Non-Koreans." *Journal of Health Care for the Poor and Underserved* 18:197–217.

Yuan, Michelle, Virginia Wake, Hyunjoo Park, and Lan Nguyen. 2009. "Conducting Cognitive Interviews with Linguistically Isolated Asian Populations." Paper presented at the International Field Directors and Technologies Conference, Delray Beach, FL, USA.

Zeldenryk, Lynne, Susan Gordon, Marion Gray, Richard Speare, Wayne Melrose, Moazzem Hossain, and Gary Williams. 2013. "Cognitive Testing of the WHOQOL-BREF Bangladesh Tool in a Northern Rural Bangladeshi Population with Lymphatic Filariasis." *Quality of Life Research* 22:1917–1926.