

Crossnational Survey Research: The Challenge and the Promise¹

Tom W. Smith

*National Opinion Research Center
University of Chicago*

Introduction

The General Social Survey (GSS) at the National Opinion Research Center/University of Chicago has been engaged in crossnational research for 22 years. This research effort began as a collaboration between the GSS and the newly organized ALLBUS, a similar program in Germany at the Zentrum für Umfragen, Methoden, und Analysen. This collaboration then expanded when the United States and Germany joined with Britain and Australia to form the International Social Survey Program (ISSP) in 1984. The ISSP has conducted a crossnational study each year since 1985 and has grown from 4 to 39 members. It now includes the founding four members plus Austria, Bangladesh, Brazil, Bulgaria, Canada, Chile, Cyprus, Czech Republic, Denmark, Finland, Flanders, France, Hungary, Ireland, Israel, Italy, Japan, Korea (South), Latvia, Mexico, the Netherlands, New Zealand, Norway, the Philippines, Poland, Por-

tugal, Russia, Slovakia, Slovenia, South Africa, Spain, Sweden, Switzerland, Taiwan, and Venezuela. From 1985 through 2000 there have been 285 surveys with over 392,000 respondents.

An essential goal of crossnational survey research is to construct questionnaires that are functionally equivalent across populations. Questions need not only be reliable and valid, but must also have comparable reliability and validity across nations. Yet the very differences in language, culture, and structure that make crossnational research so analytically valuable seriously hinder achieving measurement equivalency. Only by dealing with these challenges, in addition to the usual instrument design issues, can scientifically credible crossnational survey instruments emerge. Even ISSP's long and extensive experience in crossnational survey research does not make it an easy task.

This article discusses some of what the ISSP has learned about doing crossnational survey research and considers: (1) the development of equivalent questions in surveys, (2) measurement effects in general and variable error structures across nations, and (3) steps to enhance validity and comparability in crossnational surveys, including the form of source questions, translation procedures, and item development and pretesting.

Tom W. Smith is Director of the General Social Survey at the National Opinion Research Center, University of Chicago. He is cofounder of the International Social Survey Program and was its Secretary General from 1997 to 2003.



Asking Questions

Question wordings and their translation are “the weakest link” in achieving crossnational equivalence (Kumata and Schramm, 1956). Questions have two parts — the body of the item in which the substance and the stimulus are presented and the response scale in which the answers are recorded.

The first consideration in translating a question is the substantive meaning and conceptual focus of the question. The goal is achieving functional equivalence across versions of the questionnaire. One needs an optimal translation in which the best words are used to cover the same concepts as in the original version. Or, the more desirable situation occurs when two or more versions are developed simultaneously using words in each language that are the closest possible matches.² But even an optimal translation may not produce equivalency.

Even cognates between closely related languages can substantially differ. For example, the concept “equality/égalité” is understood differently in America, English-speaking Canada, and French-speaking Canada (Cloutier, 1976). Likewise, for Spanish-speaking immigrants in the United States, the meaning of the word “educación” includes social skills of proper behavior that are essentially missing from the more academic meaning of “education” in English (Greenfield, 1997).

Another problem occurs when a concept is represented by a word in one language but there is no corresponding word in another language. For example, a study of Turkish peasants (Frey, 1963) concluded that “there was no nationally understood word, familiar to all peasants, for such concepts as ‘problem,’ ‘prestige,’ and ‘loyalty’...” Similarly, the Japanese concept of “giri” [having to do with duty, honor, and social obligation] has no “linguistic, operational, or conceptual corollary in Western cultures” (Sasaki, 1995).

Besides language incompatibility, differences in conditions and structures also hinder achieving functional equivalence. First, situational differences can interact with words that may have equivalent literal meaning to

produce questions with different social implications. As Bollen et al. (1993) note:

Consider the young woman who has reached her family size goal. In the United States, if you ask such a woman whether it would be a problem if she were to get pregnant, she is likely to say yes. In Costa Rica, she may say no. This is because in Costa Rica, such a question may be perceived as a veiled inquiry about the likely use of abortion rather than a measure of commitment to a family size goal.

Also, structural differences mean that equivalent objects may not exist or that terms used to describe one object in one country describe something else in another country. For example, the American food-stamp program has no close equivalent in most other countries. In other cases, questions must ask not about the literal translation, but the functionally equivalent object. For example, most questions asking about the American president would inquire about the German chancellor and the Israeli prime minister and not the German or Israeli president.

Variations in conditions and structures mean that what one asks about and how one asks about objects differs across societies. This applies to behaviors and demographics as well as to attitudinal and psychological measures. For example, a study in Mali added to the standard American occupational classifications of how jobs relate to data, people, and things a fourth dimension on relating to animals (Schooler et al., 1998). Similarly, items about spouses have to allow for multiple mates in Islamic and most African societies.

Demographics can be among the least compatible of variables. Some demographics must use country-specific terms for both questions and answers. For example, region of residence uses country-specific units (e.g., “states” in the United States, “provinces” in Canada, “länder” in Germany) and of course the answers are unique geographic localities. Likewise, voting and party preference must refer to country-specific candidates and political parties.

Then there are demographics that might be asked in either country-specific or generic, cross-country terms.

For example, a generic education question might ask, "How many years of schooling have you completed?" A country-specific approach might ask about the highest degree obtained, the type of school attended, and/or the examination passed. The ISSP, for example, follows the latter course, judging that getting precise country-specific information on education is important. The former produces a simple, superficially-equivalent measure, but lumps together people who have been educated in completely different educational tracks within a country. But the latter has to struggle with comparing unique, country-specific, educational categories across nations.

With problems of linguistic and structural equivalence added to the already notable monolingual challenge of creating valid measures, the need for multiple indicators is greatly reinforced. Even with the most careful of translations, it is difficult to compare the distributions of two questions that

employ abstract concepts and subjective response categories (Smith, 1988). It is doubtful

that responses to the ISSP item, "If you were to consider your life these days, how happy or unhappy would you say you are, on the whole...very happy, fairly happy, not very happy, not at all happy, can't choose?" would be equivalent across languages. In all likelihood, the closest linguistic equivalent to "happy" will differ from the English concept in various ways, perhaps conveying different connotations and tapping other related dimensions (e.g., satisfaction), but at a minimum probably expressing a different level of intensity. Similarly, the adjectives "very," "fairly" and "not at all" are unlikely to have precise equivalents. Even in the situation in which the English adjective "very" is consistently (and correctly) translated into the French "très," it is not known whether the strength of these two words is sufficiently identical to cut the underlying continuum of happiness at the same point.

The increased need for multiple indicators in crossnational research can be illustrated by a scheme used

to compare the French and the Americans on psychological well-being:

1. A measure of general happiness
2. A measure of overall satisfaction
3. A scale of domain-specific satisfaction items

Franco-American comparisons on any one of these would be suspect because of possible language ambiguities. Even the multi-item measure of domain-specific satisfaction would be insufficient since all items utilize the term "satisfaction" and any nonequivalence would be compounded across items. Nor would the combination of the domain-specific and overall satisfaction measures solve the problem since any disparity in the meaning of "satisfaction" across languages would be perpetuated. However, switching to asking about how "happy/heureux" one is adds a question that is distinct from the satisfaction

item and avoids correlated, linguistic error from repeated terms.

"An essential goal of crossnational survey research is to construct questionnaires that are functionally equivalent across populations. Questions need not only be reliable and valid, but must have comparable reliability and validity across nations."

If linguistically-distinct measures are used,

then it is possible to get unambiguous results if the results across items are consistent (e.g., the French leading/trailing the Americans on all measures). With one measure, it is impossible to know whether any measured differences (or nondifferences) are societal or merely linguistic. With two measures, a consistent pattern on both items establishes a clear finding, but if the measures disagree it is possible that one is societal and the other linguistic and there is no basis for identifying which is which. What is desirable is to have three linguistically-distinct measures of the same construct.³ If all three agree, one has a clear, robust finding. If two agree and the third shows a different pattern, one has to be more cautious with the results, but there is at least a "preponderance of evidence" toward one substantive interpretation of the crossnational differences. If all three results disagree (positive, negative, and no difference), then no firm evidence about crossnational differences exists and much further developmental work is needed. A similar approach is called "triangulation" (Van de Vijver and Leung, 1997).

Another consideration in translating questions is the answer-recording aspect of questions. Several approaches have been offered to increase equivalency across questions (and ultimately the answers to the questions) in crossnational research.

One proposed solution is using nonverbal scales, either numeric, like feeling thermometers and magnitude measurement scaling, or visual, like ladders or Allenbach's happy/unhappy faces. While such scales are potentially useful, there is little evidence that they improve measurement or make it more comparable across countries.

A second proposal is to use only simple response scales such as dichotomies. This approach does simplify measurement, but at the cost of losing much precision by capturing only direction and not extremity.

A third proposed solution is to calibrate response scales by measuring and standardizing the strength of the labels used. One procedure has respondents rate the strength of terms as a point on a continuum. For example, one can rate each term on a numerical scale (using a 10- or 21-point scale) (Smith, 1997). This measures absolute strength and the distance between terms and thus facilitates the creation of equal interval scales.

The direct-rating approach was used in a pilot study of terms employed in response scales in Germany and the United States (Smith, 1997; Mohler, Smith, and Harkness, 1998) with very promising results. Many response terms were highly equivalent in Germany and the United States, but some notable systematic differences also appeared.⁴

In addition to the technical challenges that this approach demands, a major drawback is that separate methodological studies are needed in each country and language to establish the calibration. This obviously is something that every crossnational study cannot undertake. However, in theory, once calibrations are determined they could be used by other studies without extra data collection needed. Moreover, since the same response scales are used across many different substantive questions, a small number of carefully calibrated response scales could be used in many questions.

Measurement Effects

Crossnational comparability is also difficult to achieve because of differences in measurement effects. The special danger in crossnational surveys is that various error components correlate with the nation being studied and therefore observed differences could represent differences in response effects rather than in substance. Saris' work (1998) across 13 cultural areas found that "even if the same method is used, one can get different results due to differences in the error structure in different countries." Important crossnational sources of measurement variation include effects related to social desirability, acquiescence, extremity, no opinion, middle options, context/order, and mode.

Unfortunately, research on how these effects vary across countries is limited. While some examples of variable effects have been documented (Javeline, 1999; Smith, 1996; Smith, 2000; Van Herk, 2000), this does not mean that response effects are always or even typically dif-

International Social Survey Program (ISSP) holdings at ICPSR

<i>Title</i>	<i>Study No.</i>	<i>Title</i>	<i>Study No.</i>
International Social Survey Program (ISSP), 1985–2000	3881	Role of Government I and II, Cumulative File, 1985–1986, 1990	6233
Environment, 1993	6640	Role of Government II, 1990	6010
Family and Changing Gender Roles II, 1994	6914	Role of Government III, 1996	2808
Family and Changing Sex Roles, 1988	9591	Role of Government, 1985–1986	8909
National Identity, 1995	2474	Social Inequality III, 1999	3467
Networks and Support Systems, 1986	9205	Social Inequality, 1987	9383
Religion, 1991	6234	Social Inequality, 1992	6493
Religion II, 1998	3065	Work Orientations, 1989	9784

ferent across groups and societies. A number of consistent results have also been documented. For example, some social-desirability effects have been shown to be similar in Canada, the Netherlands, and the United States: telephone surveys produce lower quality data in the same countries, and forbid/allow question variations have like effects in both Germany and the United States (Hippler and Schwarz, 1986; Scherpenzeel and Saris, 1997). But variable measurement effects remain a serious concern and one that researchers must continually look out for. Crossnational surveys need to use the best possible techniques to reduce each of these effects and thus minimize the likelihood of variable effects across countries.

Enhancing Question Comparability

Various steps can be taken to enhance equivalence and therefore achieve valid crossnational research. These include: (1) crossnational cooperation over study design and questionnaire content, (2) adopting a master questionnaire using question forms more conducive to reliable measurement and suitable for translation, (3) considering both emic and etic items (see the discussion on page 7 on this topic), (4) following optimal translation procedures, (5) careful item development and pretesting, and (6) thorough documentation of survey practices.

Make Crossnational Research Collaborative

Research imperialism, in which a research team from one culture develops a project and instrument and rigidly imposes it on other societies, should be avoided. As Van de Vijver and Leung (1997) have observed, “Many studies have been exported from the West to nonwestern countries and some of the issues examined in these studies are of little relevance to nonwestern cultures.”

Instead, a collaborative, multinational approach should be followed. For example, Sanders (1994) noted:

One of its [the ISSP's] greatest strengths is that a country can only be incorporated in the survey if a team of researchers from that country are available... to ensure that the translation of the core questions can be achieved without significantly altering their mean-

ing. The potential problem of crossnational variation in meaning is accordingly minimized.

Question Form and Content

The first step in developing a questionnaire is to formulate items that make translations easier and avoid problematic constructions. Brislin (1986) in particular has 12 guidelines for making items more translatable.

Other general rules about how to formulate questions have usually been developed only within monocultural contexts, but many are applicable across countries (e.g., Converse and Presser, 1986; Fowler, 1995; Sudman and Bradburn, 1982; Van der Zouwen, 2000). These include such guidelines as avoiding vague and ambiguous wordings, double-barrelled questions, and hypothetical items.

In addition to following standard rules on constructing items, one should follow the rule that “more is better.” As discussed above, multiple indicators both enhance scale reliability and reduce linguistic artifacts.

Emic and Etic Questions

Etic questions are items with a shared meaning and equivalence across cultures and emic questions are items of relevance to some subset of the cultures under study. Suppose that one wanted crossnational data on political participation in general and contacting government officials in particular. In the United States, items on displaying bumper stickers, visiting candidate Web sites, and e-mailing public officials would be relevant. In most developing countries, these would be rare to meaningless. Conversely, an item on asking a village elder to intervene with the government might be important in developing societies, but have little relevance in developed nations. In such circumstances, solutions include (1) using general questions that cover the country-specific activities within broader items, (2) asking people in each nation both the relevant and irrelevant participation items, or (3) asking a core set of common items (e.g., voting in local and national elections, talking to friends about politics), plus separate lists of country-specific political behaviors.⁵

Using general items is perhaps the least appropriate since the necessary loss of detail is usually a heavy price to pay and general items may be too vague and sweeping.

The relevant/irrelevant approach makes sense if the number of low relevancy items is not too great and they are not so irrelevant that they do not make sense or are otherwise inappropriate. For example, the ISSP successfully used this technique in its study of environmental change where items on personal car use were asked in all countries, even though ownership levels were quite low in a few countries.

The emic/etic approach is useful if the common core is adequate for direct comparisons. For example, a study of obedience to authority in the United States and Poland had five common items plus three country-specific items in Poland and four in the United States (Miller, Slumczynski, and Schoenberg, 1981). This allows both direct crossnational comparisons as well as more valid measurement of the construct within countries (and presumably better measurement of how constructs worked in models). In effect, the emic/etic approach indicates that sometimes one needs to do things differently in order to do them equivalently.

Translation Procedures

Translation is often wrongly seen as a mere technical step rather than as being central to the scientific process of designing valid crossnational questions. Translation must be an integral part of the study design and not an isolated appendage. As Pasick and colleagues (1996) describe the designing of a multilingual study, translation is an integrated and interactive part of an eight-step process. These involve (1) conceptual development of topics, (2) an inventory of existing items, (3) development of new questions, (4) question assessment through translation, (5) construction of full, draft questionnaires, (6) concurrent pretesting across all languages, (7) item revision, and (8) final pretesting and revisions. What is essential in this process is that translation be part of (a) a larger process of item development and testing, and (b) a multistage, interactive process where changes in source and target language wordings occur at various points in the design process.

Achieving optimal translation begins at the design stage. Crossnational instruments should be designed by multinational teams of researchers who are sensitive to translation issues and take them into consideration during the design and development stages. They need to consider how each concept can be measured in each language and society under study. Specifically, they should practice decentering (Harkness and Schoua-Glusberg, 1998). Decentering is the process by which questions are formulated so they are not anchored in one language, but fit equally well in all applicable languages.⁶ Of course, the problems of translation in general and decentering in particular multiply as the number of languages involved increases and as the linguistic and cultural differences between languages widen.

There are various techniques for carrying out translations. First, there is the translation-on-the-fly approach under which multilingual interviewers do their own translations when respondents do not understand the source language. This approach obviously lacks standardization and quality control.

Second, there is the single-translator, single-translation approach. This method has never been formally recommended, but it is frequently used because it is quick, easy, and inexpensive.

Third, there is the back-translation technique under which (1) questions in the source language are translated to the target language by one translator, (2) then the translation is retranslated back into the source language by a second translator, (3) the researchers then compare the two source language questionnaires, and (4) when notable differences in the two appear, they work with one or both of the translators to adjust the problematic questions of the target language. This is probably the most frequently recommended translation method (Brislin, 1970 and 1986; Harkness, 1999). A limitation of this technique is that no direct assessment is made of the adequacy of the target language questions.

Fourth, there is the parallel-translation approach under which (1) questions in the source language are translated independently by two translators into the target language,

(2) the two translations are then compared, and (3) when found to differ appreciably, the two translators meet with those who developed the source language questions to figure out the reason for the divergent translations (Bullinger, 1995). As in back translation, this is a two-translations, two-translators approach, but with more emphasis on optimizing wording in the target language. It also can be done more quickly than back translation since the two translations can be done simultaneously rather than sequentially.

Finally, there is the committee-translation approach under which a team of translators and researchers discuss the meaning of items in the source language, possible translations in the target language, and the adequacy of the translations in the target language relating to such matters as level of complexity and naturalness, as well as meaning. This approach may use parallel translation with different members of the team producing independent translations of items, or the team may work on a translation simulta-

neously and interactively. This approach maximizes interaction between translators and between translators and

other members of the research team. It also places the greatest emphasis on writing good questions and not just on merely translating words.

While careful translation procedures are essential for developing equivalent items, they are not sufficient alone. Quantitative methods should evaluate the results of the qualitative translation procedures. Several approaches for the quantitative assessment of items and translations exist. First, there is the direct evaluation of items. For example, Bullinger (1995) describes a study in which two raters independently judged the difficulty of the wordings in the source language, then two other raters evaluated the quality of the translated items, and finally two more raters assessed the back-translated items. This allowed both

qualitative and quantitative evaluation of the translations as well as inter-rater reliability checks on the quantitative ratings. Second, quantitative ratings of the terms used in response options can determine whether scale points are equivalent. Third, various statistical tests can assess the comparability of crossnational results. While usually applied at the analysis phase, they can and should be employed at the development stage. Item-response theory (IRT) and confirmatory or exploratory factor analysis have been used for this purpose (Ellis, Minsel, and Becker, 1989; MacIntosh, 1998; Ryan et al., 1999). Finally, these quantitative evaluation approaches can be combined.

The various quantitative techniques should be used hand-in-hand with qualitative techniques. For example, in the German-American study of response options (Mohler, Smith, and Harkness, 1998), equivalent English and German terms for answer scales were developed by translators and then respondents rated the strength of the terms on the underlying dimensions (agreement/dis-

agreement and importance). In most cases, the mean ratings of the German and English terms were the same, thereby validating translation equivalency.

“Translation is often wrongly seen as a mere technical step rather than as being central to the scientific process of designing valid crossnational questions. Translation must be an integral part of the study design and not an isolated appendage.”

Also, it has been proposed that translation equivalence can be established by administering items in two languages to bilingual respondents. However, this approach is problematic because bilinguals understand and process language differently than monolinguals do (Blais and Gidengil, 1993; Ellis, Minsel, and Becker, 1989). Despite this serious impediment, useful evaluations can be gained by looking at how results compare within societies, but across languages.

Achieving item and scale equivalency is a challenging task and optimal translations are essential for reaching this goal. Researchers should (1) make translations an integrated part of the development of studies, (2) utilize the best approaches, such as committee and combined

translation, and (3) use quantitative methods to assess the translations.

Pretesting and Related Questionnaire Development Work

While pretesting and piloting are important in monocultural surveys, their value greatly increases cross-nationally. Developmental work must establish that the items and scales meet acceptable technical standards (e.g., of comprehension, reliability, and validity) in each country and are comparable across countries. Moreover, the pretesting should be “a team effort with multiple disciplines and preferably multiple cultures represented” (Pasick et al., 1996).

Useful developmental and pretesting procedures include the following: (1) cognitive interviews using such protocols as think-alouds, in which respondents verbalize their mental processing of questions, and computer-assisted concurrent evaluations (Bolton and Bronkhorst, 1996; Gerber and Wellans, 1997; Gerber and Wellans, 1998; Johnson et al., 1997; Krosnick, 1999; Pruefer and Rexroth, 1996; Tourangeau, Rips, and Rasinski, 2000), (2) behavioral coding with the interviewer-respondent exchanges recorded, coded in detail, and then analyzed (Fowler and Cannell, 1996; Pruefer and Rexroth, 1996; Krosnick, 1999), and (3) conventional pretesting, including the use of probing (Converse and Presser, 1986; Fowler, 1995; Hudler and Richter, 2001).

General rules about pretesting in crossnational research include: (1) the best pretesting procedures must be carried out across countries and languages with results evaluated by researchers expert in (a) the cultures and languages being investigated, (b) the substantive domains being studied, and (c) survey-research methodology; (2) the pretesting and translating must be integrated and interactive processes; and (3) the developmental process takes much more time and effort than for single-country, monolingual studies and usually involves multiple rounds of pretesting.

Documentation

As Jowell (1998) has observed, good documentation and “detailed methodological reports about each participating nation’s procedures, methods, and success rates...” are essential. However, as Hermalin, Entwisle, and Myers (1985) have noted, “maintenance and documentation are painstaking tasks for which little provision is made...” While all phases of each survey from sampling to data processing need to be carefully recorded, it is particularly important to include the original questionnaires used in each of the countries so they can be consulted to understand results (and particularly differences in results) across countries. The ISSP, which includes copies of original instruments in its documentation, does this. Moreover, solid documentation is more than just good practice that facilitates primary and secondary analysis. It enhances comparability from the start by forcing researchers to detail the procedures are being used in each country and how comparable they are.

Conclusion

The great challenge in crossnational survey research is that languages, social conventions, cognitive abilities, and response styles all vary across societies. To obtain valid, equivalent measurement across cultures, measurement error from these sources must be minimized and equalized so that valid, reliable, and consistent substantive information emerges. Achieving this is difficult. The task of obtaining crossnational comparability is so complex and challenging that more effort is needed at all stages, from conceptualizing the research question, to instrument development, to survey analysis. But the benefits of crossnational research fully merit the extra efforts. As the Working Group on the Outlook for Comparative International Social Science Research has noted, “A range of research previously conceived of as ‘domestic,’ or as concerned with analytical propositions assumed invariant across national boundaries, clearly needs to be reconceptualized in the light of recent comparative/international findings.” Unless a comparative perspective is successfully adopted, “models and theories will continue to be ‘domestic’ while the phenomena being explained clearly are not” (Luce, Smelser, and Gerstein, 1989).

Notes

¹ Sections of this article draw upon Tom W. Smith, "Developing Comparable Questions in Cross-National Surveys," in *Cross-Cultural Survey Methods*, edited by Janet Harkness, Fons van de Vijver, and Peter Ph. Mohler. London: WileyEurope, 2002.

² See the "Translation Procedures" section.

³ This does not refer to three, single-item measures, but three linguistically distinct items or scales. For example, domain-specific satisfaction measures usually cover many different areas (e.g. job, finances, family, health, etc.).

⁴ For similar findings also based on a German/American study, see Bullinger, 1995.

⁵ However, even identical actions — e.g., voting in the last national election — may not be equivalent. In some countries, voting is legally mandatory, so it is not a meaningful measure of voluntary political activity. In other countries, elections are meaningless charades, so voting is not a meaningful measure of participating in a democracy or of making a political choice.

⁶ Decentering is not possible when a well-established scale developed in one language is being replicated across countries, but should be employed whenever new items and scales are being designed for a multilingual study.

References

- Blais, Andre, and Elisabeth Gidengil, "Things Are Not Always What They Seem: French-English Differences and the Problem of Measurement Equivalence," *Canadian Journal of Political Science* 26 (Sept., 1993), 541–555.
- Bollen, Kenneth A., Barbara Entwisle, and Arthur S. Alderson, "Macrocomparative Research Methods," *Annual Review of Sociology* 19 (1993), 321–351.
- Bolton, Ruth N., and Tina M. Bronkhorst, "Questionnaire Pretesting: Computer-Assisted Coding of Concurrent Protocols," in *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*, edited by Norbert Schwarz and Seymour Sudman. San Francisco: Jossey-Bass, 1996.
- Brislin, R.W., "Back-Translation for Cross-Cultural Research," *Journal of Cross-Cultural Research* 1 (1970), 185–216.
- Brislin, R.W., "The Wording and Translation of Research Instruments," in *Field Methods in Cross-Cultural Research*, edited by W.J. Lonner and J.W. Berry. Newbury Park, CA: Sage, 1986.
- Bullinger, Monika, "German Translation and Psychometric Testing of the SF-36 Health Survey: Preliminary Results from the IQOLA Project," *Social Science Medicine* 41 (1995), 1359–1366.
- Cloutier, Edourd, "Les Conceptions Americaine, Canadienne-Anglaise, et Canadienne-Francaise l'idée d'égalité," *Canadian Journal of Political Science* 9 (1976), 581–604.
- Converse, Jean M., and Stanley Presser, *Survey Questions: Handcrafting the Standardized Questionnaire*. Beverly Hills: Sage, 1986.
- Ellis, Barbara B., Beate Minsel, and Peter Becker, "Evaluation of Attitude Survey Translations: An Investigation Using Item Response Theory," *International Journal of Psychology* 24 (1989), 665–684.
- Fowler, Floyd J. Jr., *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage, 1995.
- Fowler, Floyd J. Jr., *Survey Research Methods*. 2nd edition. Newbury Park, CA: Sage, 1993.
- Fowler, Floyd J. Jr., and Charles F. Cannell, "Using Behavioral Coding to Identify Cognitive Problems with Survey Questions," in *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*, edited by Norbert Schwarz and Seymour Sudman. San Francisco: Jossey-Bass, 1996.
- Frey, F. W., "Survey Peasant Attitudes in Turkey," *Public Opinion Quarterly* 27 (1963), 335–355.
- Gerber, Eleanor R., and Tracy R. Wellens, "Perspectives on Pretesting: 'Cognition' in the Cognitive Interview?" *BMS* 55 (June 1997), 18–39.
- Greenfield, Patricia M., "You Can't Take It With You: Why Ability Assessments Don't Cross Cultures," *American Psychologist* 52 (Oct. 1997), 1115–1124.
- Harkness, Janet A., "Questionnaire Development, Adaption, and Assessment for the ESS," Paper presented to the International Conference on Quality in Official Statistics, May 2001.
- Harkness, Janet A., and Alisu Schoua-Glusberg, "Questionnaires in Translation," in *Cross-Cultural Survey Equivalence*, edited by Janet Harkness. ZUMA-Nachrichten Spezial No. 3. Mannheim: ZUMA, 1998.

- Hermalin, Albert I., Barbara Entwisle, and Lora G. Myers, "Some Lessons from the Attempt to Retrieve Early KAP and Fertility Surveys," *Population Index* 51 (Summer 1985), 194–208.
- Hippler, Hans-Jurgen, and Norbert Schwarz, "Not Forbidding Isn't Allowing: The Cognitive Basis of the Forbid-Allow Asymmetry," *Public Opinion Quarterly* 50 (Spring 1986), 87–96.
- Hudler, Michaela, and Rudolf Richter, "Theoretical and Methodological Concepts for Future Research and Documentation on Social Reporting in Cross-Sectional Surveys," EuReporting Working Paper No. 18. Vienna: Paul Lazarsfeld-Gesellschaft für Sozialforschung, 2001.
- Javeline, Debra, "Response Effects in Polite Cultures: A Test of Acquiescence in Kazakhstan," *Public Opinion Quarterly* 63 (Spring 1999), 1–28.
- Johnson, Timothy P., et al., "Social Cognition and Responses to Survey Questions Among Culturally Diverse Populations," in *Survey Measurement and Process Control*, edited by Lars Lyberg, et al. New York: John Wiley & Sons, 1997.
- Krosnick, Jon A., "Survey Research," *Annual Review of Psychology* 50 (1999), 537–567.
- Kumata, H., and W. Schramm, "A Pilot Study of Cross-Cultural Meaning," *Public Opinion Quarterly* 20 (1956), 229–238.
- Luce, R. Duncan, Neil J. Smelser, and Dean R. Gerstein, eds. *Leading Edges in Social and Behavioral Science*. New York: Russell Sage, 1989.
- MacIntosh, Randall, "A Confirmatory Factor Analysis of the Affect Balance Scale in 38 Nations: A Research Note," *Social Psychology Quarterly* 61 (March 1998), 83–91.
- Miller, Joanne, Kazimierz M. Slomczynski, and Ronald Schoenberg, "Assessing Comparability of Measurement in Cross-National Sociocultural Settings," *Social Psychology Quarterly* 44 (Sept. 1981), 178–191.
- Mohler, Peter Ph., Tom W. Smith, and Janet A. Harkness, "Respondent's Ratings of Expressions from Response Scales: A Two-Country, Two-Language Investigation on Equivalence and Translation," in *Nachrichten Spezial, Cross-Cultural Survey Equivalence* 3 (1998), edited by Janet A. Harkness. Mannheim: ZUMA, 1998.
- Pasick, Rena J., et al., "Problems and Progress in Translation of Health Survey Questions: The Pathways Experience," *Health Education Quarterly* 23 (Dec. 1996), 28–40.
- Pruefer, Jostein, and Margit Rexroth, "Verfahren zur Evaluation von Survey-Fragen: Ein Überblick." ZUMA-Arbeitsbericht 95/5, 1996.
- Ryan, Ann Marie, David Chan, Robert E. Ployhart, and L. Allen Slade, "Employee Attitude Surveys in a Multinational Organization: Considering Language and Culture in Assessing Measurement Equivalence," *Personnel Psychology* 52 (1999), 37–58.
- Sanders, David, "Methodological Considerations in Comparative Cross-National Research," *International Social Science Journal* 46 (Dec. 1994), 513–521.
- Saris, Willem E., "The Effects of Measurement Error in Cross-Cultural Research," in *Nachrichten Spezial, Cross-Cultural Survey Equivalence* 3, edited by Janet A. Harkness. Mannheim: ZUMA, 1998.
- Sasaki, Masamichi, "Research Design of Cross-National Attitude Surveys," *Behaviormetrika* 22 (Jan. 1995), 99–114.
- Scherpenzeel, Annette C., and Willem E. Saris, "The Validity and Reliability of Survey Questions: A Meta-Analysis of MTMM Studies," *Sociological Methods and Research* 25 (Feb. 1997), 341–383.
- Schooler, Carmi, et al., "Conducting a Complex Sociological Survey in Rural Mali: Three Points of View," *American Behavioral Scientist* 42 (Oct. 1998), 252–275.
- Smith, Tom W., "Developing Comparable Questions in Cross-National Surveys," in *Cross-Cultural Survey Methods*, edited by Janet Harkness, Fons van de Vijver, and Peter Ph. Mohler. London: WileyEurope, 2002.
- Smith, Tom W., "Environmental and Scientific Knowledge Around the World," GSS Cross-National Report No. 16. Chicago: NORC, 1996.
- Smith, Tom W., "Improving Cross-National Survey Response by Measuring the Intensity of Response Categories," GSS Cross-National Report No. 17. Chicago: NORC, 1997.
- Sudman, Seymour, and Norman M. Bradburn, *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass, 1982.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski, *The Psychology of Survey Response*. Cambridge: Cambridge University Press, 2000.
- van der Zouwen, Johannes, "An Assessment of the Difficulty of Questions Used in the ISSP-Questionnaires, the Clarity of Their Wording, and the Comparability of the Responses," *ZA-Information* No. 46 (2000), 96–114.
- van de Vijver, Fons, and Kwork Leung, *Methods and Data Analysis for Cross-Cultural Research*. Thousand Oaks, CA: Sage, 1997.
- van Herk, Hester, *Equivalence in a Cross-National Context: Methodological and Empirical Issues in Marketing Research*. Published Ph.D. dissertation, Catholic University, Brabant, 2000.

ANNOUNCEMENTS

NICHHD Awards Grant for Disclosure Risk Analysis Project

Researchers from ICPSR and the Survey Research Center (SRC) of the Institute for Social Research were recently awarded a National Institute on Child Health and Human Development (NICHD) grant on Human Subject Protection and Disclosure Risk Analysis. ICPSR Director Myron Gutmann is the overall principal investigator for the project, which involves four individual research efforts.

Project 1, headed by Eleanor Singer of the SRC is "Informed Consent and Perceptions of Risk and Harm in Survey Participation." Singer and SRC co-investigators Fred Conrad, Mick Couper, and Bob Groves will study the level of risk of disclosure that the public is willing to accept; whether disclosure of some kinds of information is considered more harmful than others; whether some data intruders are perceived as more harmful than others; whether people perceive the relationship of expected risk of harm versus magnitude of harm and risk of disclosure in a manner consistent with the mathematical probability of such occurrences; and how researchers can accurately inform participants of the risks without unnecessarily deterring them from participation.

T.E. Raghunathan (SRC) will lead Project 2, entitled "Estimation of Disclosure Risk and Statistical Methods for Disclosure Limitation." Co-investigators on this project are Ben Hansen, Rod Little, and Richard Valliant, also from the SRC. Their objectives include (1) assessment of the risk of disclosure using test-bed national probability surveys covering diverse topics; (2) development and evaluation of new methods to prevent disclosure; and (3) development of methods for constructing coarsened, perturbed, or synthetic versions of sensitive variables in public-use datasets.

Project 3 is headed by JoAnne McFarland O'Rourke at ICPSR, with Myron Gutmann as co-investigator and Corey Colyer as Research Associate. The project, "Best Practices and Tools for the Social Sciences," will develop best practices for disclosure limitation (1) by reviewing the literature on disclosure, (2) surveying the principal investigators and others involved in disclosure decisions for a sample of studies funded by NIH and NSF, (3) using these results and relevant findings from Projects 1 and 2 to define best practices for different types of data, and (4) designing tools that incorporate best practices.

James McNally (ICPSR) leads the fourth and final project, "Resources for the Secure Dissemination of Human Subject Data," with Myron Gutmann as co-investigator. This project serves as a bridge to connect the first three projects and to share tools, education materials, and outcomes with the research community. This will be accomplished by (1) supporting a Web site to disseminate research findings and tools, (2) organizing a series of lectures and training seminars, and (3) maintaining a central knowledge base on disclosure risk analysis by providing links to vital information resources.

Look for Changing URLs on ICPSR Web site

ICPSR has recently upgraded some parts of the Web site software infrastructure. This means that some URLs have changed. Here is an example of the change:

old URL —

<http://www.icpsr.umich.edu:8080/ICPSR-SERIES/00035.xml>

new URL —

<http://webapp.icpsr.umich.edu/cocoon/ICPSR-SERIES/00035.xml>

This change only affects the XML files on the Web site, which includes the study descriptions, series descriptions, CD-ROM abstracts, union catalog entries, and thesaurus pages. Users who have bookmarked or linked to those pages will be affected.

This upgrade improves ICPSR's overall capabilities with regard to XML display, and it is also intended to facilitate access to the ICPSR Web site for those visitors that are behind firewalls, where outbound access to "off number" ports is blocked, but access to well-known ports is open. Previously, users who worked in secure computing environments (i.e., behind firewalls) experienced some difficulties retrieving XML files from the Web site.

Both the old and the new URLs will be active until the end of March, when we will retire the old URLs.

In Memoriam — Heinz Eulau: 1915–2004

Heinz Eulau, a long-time advocate of ICPSR and an important figure in its history, died on January 18, 2004, at his home on the Stanford University campus, where he was William Bennett Munro Professor of Political Science, Emeritus. He was 88. His wife Cleo, an adjunct clinical professor in the Department of Psychiatry and Behavioral Sciences, died shortly afterward on January 23 at the age of 80.

A tireless supporter of the Consortium and its staff, Heinz served on the ICPSR Council from 1967 through 1970, holding the position of Chair during 1968-1970. He served on Council again from 1973 through 1978 and was one of four Associate Directors during the 1980s and 1990s.

On hearing of Heinz's death, former ICPSR Director Jerome Clubb said, "The Consortium never had a better or more loyal friend than Heinz Eulau." Erik Austin, ICPSR Assistant Director, said, "For Heinz, the Consortium embodied the valuable (but rare) ideal of scholarly cooperation in pursuit of improved explanations of society and social structures. He invested large amounts of his own time in the service of this organization that he cared so much about. He will be greatly missed."

For more information about Heinz's life and his contributions to the field, please see the Stanford Report obituary at <http://news-service.stanford.edu/news/2004/january28/eulaubit-128.html>.

IASSIST 2004 Conference — *Data Futures: Building on Thirty Years of Advocacy*

The International Association for Social Science Information Services and Technology (IASSIST) annual conference will be held at the University of Wisconsin, Madison, May 25-28, 2004. This year's conference, *Data Futures: Building on Thirty Years of Advocacy*, examines new issues and trends and links them to principles that have emerged during the past 30 years.

IASSIST has been on the leading edge of data dissemination and access issues, critically examining developments in electronic delivery and privacy/confidentiality concerns. "Data advocacy" has included promoting statistical literacy among data professionals and the public, participating in the development of metadata standards for data, and working on solutions for preservation and archiving. The 2004 conference will address various aspects of data advocacy.

For more information, visit <http://dpls.dacc.wisc.edu/iassist2004/index.html>.

LEHD and DEED Datasets Available at the Michigan Census Research Data Center

The Michigan Census Research Data Center (MCRDC) is pleased to announce that LEHD and DEED datasets are now available through the research data center located in the Institute for Social Research at the University of Michigan, Ann Arbor. The MCRDC is now fielding proposals for the use of the LEHD and the DEED datasets. Both of these datasets are employer-employee matched, opening up new directions for research and analysis.

LEHD. The Longitudinal Employer-Household Dynamics (LEHD) is a partnership program between the U.S. Census and individual states that integrates state unemployment insurance (UI) with Census data, providing information to the policy-making and research communities about the dynamics of economic activity. This resource is composed of four distinct datasets: (1) LEHD Business Register Bridge, (2) LEHD Employer Characteristics File, (3) LEHD Employer Human Capital File, and (4) LEHD Employer Quarterly Workforce Indicators. These datasets allow for the integration of Census economic data with employee characteristics files.

DEED. The Decennial Employer-Employee Database (DEED) is the 1990 Decennial Census linked place of work from the 1990 Standard Statistical Establishment List, which allows for the joining of 1990 Decennial Census data to various Economic Censuses.

The Michigan Census Research Data Center (MCRDC) allows qualified researchers with approved projects to conduct research using unpublished microdata from the Census Bureau's economic and demographic programs. All MCRDC research is conducted within its secure laboratory facility located in the Institute for Social Research and must have a Census Bureau purpose.

Researchers are invited to submit proposals to use the MCRDC. Proposals are now *accepted throughout the year starting March 2004*. Researchers with projects in the MCRDC must have special sworn status with the Census Bureau. The MCRDC assists researchers in obtaining this status. Researchers from ICPSR member institutions may apply to use the seat without paying MCRDC laboratory fees.

Please refer to the MCRDC and the Census Bureau Center for Economic Studies (CES) Web pages for information about the proposal process and available data sets.

CES Web page: <http://www.ces.census.gov/ces.php/home>

MCRDC Web page: <http://www.isr.umich.edu/src/mcrdc/mcrdc.html>