The Global Survey of Physicists: Challenges and Lessons Learned from an Evolving Multilingual Multinational Survey Effort

Casey Langer Tesfaye, Rachel Ivie
American Institute of Physics, One Physics Ellipse, College Park, MD 20740

Abstract
Linguistic and cultural diversity are challenges that have amplified effects within the quantitatively driven, sensitive field of survey research, but are an important part of the future of the field. The Global Survey of Physicists, produced by the American Institute of Physics (AIP) on behalf of the International Union of Pure and Applied Physics (IUPAP) Working Group on the Status of Women in Physics provides a unique case study in multinational, multilingual and multicultural research. The survey uses nontraditional distribution methods, and it has been continually reshaped and redesigned as a result of its challenges and successes. We will address the evolving methods of the global survey, as well as its increased effectiveness and reach. We will also examine the many complications associated with the collection of reliable and valid data in a multicultural context.

Key Words: multinational, multilingual, multicultural, translation, snowball, nonprobability

1. Introduction

"Probably no field has generated more methodological advice on a smaller database with fewer results than has [cross-national] comparative sociology." William Form, 1979 (Quoted by Kohn, December 1987)

In April of 2008, Colm O'Muircheartaigh gave a talk about the past, present and future of survey research. He suggested that in order to embrace the complicated underlying diversity inherent in the communities we survey, we need to continue our evolution from our more rigid experimental roots, which favor artificially homogeneous populations, (O'Muircheartaigh, April 11, 2008). The world is diverse, culturally and linguistically, and accurately embracing this diversity will require us to look critically at the complications that diversity brings to the survey process as we have understood it (from a traditional monocultural perspective.) In order to produce valid conclusions, we need to look critically at our methodological challenges and successes and have the courage to question our conclusions. This paper is a critical examination of the methodological challenges and successes of The Global Survey of Physicists, a multinational, multilingual and multicultural survey produced and maintained by the American Institute of Physics (AIP).

1.1 About the Survey

There is no one global statistic about the prevalence of women in physics, but we do know that women are severely underrepresented. Although there has been a large and steady increase in the proportion of women in physics in the past century, the statistics that are available reliably show that across the globe, women comprise less than one fifth of the recent physics PhDs (Ivie & Guo, 2006).

As a result, gender is an important topic of concern within the physics community. Women often feel quite isolated and sometimes experience overt and covert discrimination in their pursuit of physics. Many women leave the field early, despite their love of the subject matter. And many women who work in the field of physics feel a distinct disadvantage compared to their male counterparts. The International Union of Pure and Applied Physics (IUPAP) formed a working group on women in physics in 1999. In order to describe the experiences of women in physics across the globe, the American Institute of Physics (AIP) conducted a study on behalf of the working group in conjunction with the first IUPAP International Conference of Women in Physics in 2002. The study was very well received, and in 2005 and then again in 2009, the working group set out to repeat and improve it. The purpose of the Global Survey of Physicists is to describe commonalities and differences within the growing, widespread population of female physicists.

The relationship between the IUPAP working group and the AIP survey group is unique. The working group and the AIP survey researchers are in close contact with a larger group of team leaders from each IUPAP country. The team leaders are in turn, in close contact with women physicists in their respective countries. Members of the working group, team leaders, and attendees to conferences sponsored by the working group have helped with everything from the design of the survey to its distribution.

1.2 Past Surveys

In 2002, the first Global Survey of Physics was a complicated flagship effort. Members of the working group developed questions in English, and then inserted them into the body of an e-mail for distribution. Attendees to the first IUPAP conference on women in physics forwarded the e-mail survey to their female colleagues, asking the recipients to complete the survey and then pass the e-mail on to their female colleagues in physics. When the responses proved to be of limited use, the Statistical Research Center at AIP was asked to rework the survey, and it was re-circulated. The survey effort generated responses from over 1000 women in 50 countries across the globe. The report supplemented quantitative findings heavily with qualitative observations and direct quotes.

In 2005, the American Institute of Physics conducted the Global Survey of Women for the second time, building on the previous survey. This time AIP developed a standardized web survey that both maintained continuity with the e-mail instrument distributed in 2002 and improved some of the questions. Similar to the 2002 survey, the 2005 survey was distributed through a snowball sample. An IUPAP representative sent the survey link to the 150 attendees of the Second International Conference of Women in Physics with instructions to pass the survey along to their female colleagues. This survey gathered over 1350 responses from just over 70 countries. However, the attendees participated to varying degrees, creating dramatically mixed levels of response between countries.

Because the first two surveys were not based on a representative sample, the results could not accurately represent of the global population of women in physics. The surveys were in English only, representing a barrier for many respondents. In addition, data were not collected from men, making it impossible to benchmark differences by gender. But despite these significant limitations, the surveys produced an unprecedented level of description of some of the difficulties encountered by women in physics across a great breadth of countries, as well as a glimpse into the differences between women physicists in developed and developing countries. The surveys contained valuable qualitative data

from respondents that supplemented the quantitative data. The reports generated quite a bit of enthusiasm within the physics community. The team leaders were happy to see their experiences validated by the survey and set out to repeat the survey again in 2009, in association with the 2008 meeting of the working group.

2. The 2009 Survey

2.1 Design changes

The design of the 2009 Global Survey of Physicists differs markedly from its predecessors. Although many of the core questions were carried over from past surveys, many other questions were changed or added. The American Institute of Physics began the survey design process with a thorough review of other surveys of women in physics conducted ad hoc across the globe. Questions that were repeated in multiple surveys or that we judged as particularly important or insightful were added or adapted into the Global survey. Significant input was solicited and incorporated from members of the working group, and comments and results from the 2005 survey informed survey improvements.

Language fluency and familiarity with online surveys were significant areas of concern to us in the survey design process. We worried that some respondents would have varying degrees of language comprehension when they approached the survey. We chose a paging design for the survey in an effort to minimize confusion: what was a lengthy single page scrolling survey in 2005 became a 23 module paging survey in 2009. Research in monolingual studies has repeatedly shown that skip patterns can be confusing or simply ignored by respondents (Dillman, 2007). This effect could easily be magnified with our respondents. Instead each section was divided into a module with screening questions and a more in-depth module with questions specific to each topic. The more in-depth questions were only shown to respondents who had indicated in the screening question that the subsequent module would apply to them.

Paging surveys also allow researchers to more narrowly focus on the point of drop-offs for partial survey responses (Couper, 2008). In preliminary analysis of the Global Survey, although the total number of incomplete responses was significant, drop-off levels were consistent throughout the modules. This shows that partial responses were probably more a result of the overall length of the survey than a reaction to a specific question or set of questions within it. Brevity is always important in survey research, but it is likely especially important in multilingual, multicultural survey.

Although we did not do cognitive interviewing on the Global Survey, we did circulate the web address among the IUPAP team leaders for a feedback period in order to isolate potential cultural differences before the costly translation process began. Although we received quite a bit of feedback from some team leaders, this was a very passive method of inviting review of the survey. In order to ensure a more complete review and better comparability of the results in future surveys, it will be important to involve the team leaders in a more active way. This could be accomplished remotely item by item, or by recreating the survey in a collective environment, involving active discussions about each construct and what it means to each of the team leaders.

2.2 The translation process

As part of the planning process for the 2009 survey, the IUPAP working group decided to translate the survey into 7 languages in addition to English: Spanish, French, Arabic,

German, Russian, Japanese, and Chinese. A grant from the Henry Luce Foundation helped to fund the translation. The translation process took much more time than we had budgeted and raised quite a few nuanced and complicated issues that we had not anticipated.

Although we had to sacrifice translation into a few useful additional languages, such as Farsi and Italian, we chose a professional translation service over the informal offers that we received from working group members. We hoped that this would bring better control over the process, free us from the politics that every working group brings, and standardize the translations as much as possible. Translation quality was paramount to our goal-- to have all of our questions comparable across all of our language versions and generate comparable responses. This goal was a wise, but shortsighted measure of quality control.

Before beginning the translation process, we tried to make sure that the survey was ready for a smooth transition. Brislin (1986) has an especially good summary about writing for translation (below). Many of these points dovetail with guidelines for writing survey questions.

1. use short, simple sentences of less than 16 words (one idea per sentence)
2. employ active rather than passive voice (easier for translator to identify the parts of speech)
3. repeat nouns rather than using pronouns
4. avoid metaphors and colloquialisms
5. avoid the subjunctive (other languages won't necessarily have equivalents)
6. add sentences to provide context for key ideas
7. avoid adverbs and prepositions telling "where" or "when"
8. avoid possessive forms
9. avoid vague modifiers like probably, maybe or perhaps
10. use wording familiar to the translators
11. avoid sentences that have two verbs that suggest separate actions

(Brislin, 1986)

Survey translation may seem simple. "Surveyspeak" (Harkness, 2010), or the particular language used in surveys, usually includes short, clear phrasing and minimizes excess language. Harkness & Schoua-Glusberg (1998) observed that surveys generally appear to be easy to translate, with simple vocabulary and syntax, short length, and short sentences. But this is not the full story. As Harkness (2010) notes, language is not isomorphic, and what goes in cannot equal what comes out: "A mechanistic notion of input and output is itself misleading." In fact, because surveys seem so straightforward, they are often not adequately screened for potential cultural differences, issues, problems and areas of non-congruence. But the translation process is neither simple nor straightforward, and the cultural and linguistic differences are numerous. Additionally, surveys are particularly sensitive, and even a small change in the directionality of a question or the preconceptions of the respondent toward a question, both of which are correlated with the cultural background of the respondent, could hamper the analyst's ability to make valid conclusions across language versions.

Some researchers approach these complications by supplying a great deal of supplemental documentation to the translators. The decision of which supplemental

documentation to supply to the translator is a difficult one. Harkness et al. (2003) note that "translators translate what they think they see, or what they think is intended." They advocate for supplemental texts and suggest that additional documentation and instructions are important to supply to the translators. But this is one area where traditional survey knowledge conflicts with the needs of the translation process. Schaeffer (2007) points out the well documented, significant differences that even small changes in questionnaire wording can make. Kleiner et. al (2009) report that the introduction of instructions and supplemental documentation to the translator has mixed results. The documentation decreases the linguistic integrity of the translated survey in comparison with the original survey, but it increases the semantic fidelity of the translated survey to the original survey. Kleiner et al. suggest a review of project goals when setting up the survey for the translation process. Is it more important to maintain consistency among the translated versions of the surveys? Or is the ultimate goal of the survey to collect comparable responses? If the goal is to collect comparable responses, differences in the source document and translations are not as important.

Braun (1995) describes potential sources of error with translated surveys. Questions can be poorly translated, causing respondents to read and respond to the wrong question. Well-translated questions could be interpreted or framed differently in different cultural contexts. Differences in local norms could affect question processing and answers supplied. And the assumptions of the respondent could be problematic: respondents to cross-cultural surveys assume the origin is local and contextualize under that assumption. Tom Smith (2009) rightly observed that "the very differences in language that make cross-cultural research so analytically valuable are the same that seriously hinder the achievement of equivalency."

Given these circumstances, it is essential to implement an outside review process for translations. Reviews of translations provide an excellent source of quality control, so that the final translation includes both the linguistic expertise of the translators and the subject matter expertise of the reviewers. It is particularly difficult for reviewers to anticipate cultural differences prior to translation, and many cultural and linguistic differences must be reconciled at this stage. Some translation has to be done before any feedback about generalizability is requested, because even experienced cross-cultural researchers can fail to see potential differences before the first working translation is complete (Harkness & Schoua-Glusberg, 1998).

We chose to involve our working group in the translation review process. In our case, this not only helped with quality control, but also helped the team leaders feel connected to the project. The review process is quite difficult operationally. In order to properly review the translation, the reviewer must see both the source document and the translation that they are reviewing, side by side. In the case of html scripting, the interface must have some html functionality (so that, for example, a reviewer can read the choices on a pull-down menu without seeing the html code that generated it) as well as an ability to make and track changes. This is further complicated when there are multiple reviewers looking at a single translation. In our case, this process lasted much longer than the rest of the translation process and the survey development process, and it almost doubled the total cost of the translation project. But the review process is an important aspect of quality control. The reviewers were more familiar with the common words for subject-matter concepts that were more obscure for the translators. In some cases, our reviewers made large changes, such as changing the gender of the survey instrument. And in some cases, the language of the translation was so loyal to the English survey,

exhibiting such a high level of syntactic interference, that it did not make sense to reviewers who were not fluent in English.

Despite pretesting and research, the review process exposes significant cultural differences. In the Global Survey of Physicists, these differences included the use of habilitation degrees (degrees earned after doctoral degrees and necessary prerequisites for academic positions in some European countries), punctuation differences between Europe and North America that made it impossible for us to use decimals across linguistic diasporas, and difficulties with respect to our questions about marriage. For example, representatives from some Scandinavian countries had insisted that long-term partnerships be valued equally with marriages on the survey, and the inclusion of marriage and partnerships together deeply offended reviewers in parts of the Middle East. There were also difficulties in translating concepts, such as postdoctoral fellowships, that were not meaningful in some of our target countries, but essential in others. The propagation of differences at this point in the translation process underscores the difficulty involved in producing equivalent translations.

Harkness & Schoua-Glusberg (1998) suggest that an ideal translation environment includes researchers, questionnaire designers, target-language implementers and the translators themselves. This kind of environment can be very difficult to create, but it is certainly advantageous. They also suggest that the translation and review process should involve both bilinguals and monolinguals. In this round of the Global Survey of Physicists, we ultimately included all of the proper elements necessary for survey translation. But in future rounds, these influences must become more intentional in order to be maximally beneficial.

In fact, Heath et al. (2009) advocate for a de-centering process before the translation has even begun. Equivalence problems can stem from a lack of common concepts, a common concept interpreted differently, or a poor translation. A de-centering process is a process of constructing a common linguistic and cultural basis before the translation begins. Rather than creating an English source document and imposing it on the translators, common concepts are developed by a group of interested and invested parties, from which to create source documents in each language. This can be done through a series of back translations, which are repeated until a common basis is formed. Or it can be done in committee, where bilingual and monolingual subject-matter experts, survey experts and translators gather to create working documents. Although on its surface, Heath's suggestion may appear costly and difficult to coordinate, the amount of time and expense that it could potentially save make it a very sensible solution. On the other hand, experienced veterans of the de-centering process often worry about the resulting lack of cultural specificity, particularly in multilingual surveys. De-centering is thought to be most effective for bilingual surveys.

2.3 Distribution

The 2009 Global Survey of Physics is being distributed as a random sample in some populations, but there is also a snowball sample component. In October 2009 we mailed the survey link to IUPAP team leaders located across almost 90 countries with instructions to distribute the survey among their colleagues. We added a plea to the final module, the "thank you page," which asks respondents to pass the survey among their colleagues. The rationale for the snowball distribution in some countries was that there was no viable alternative, e.g., no list of physicists from which to draw a sample. In countries that used a random sample, respondents from the sample were encouraged to

pass along the survey to increase the number of women. The number of women physicists in some countries is very small, and we could not afford to miss any. Those who were not part of a random sample will not be included in any weighted analysis of random samples, but they will be included in the general analysis. As of the writing of this paper, the 2009 Global Survey of Physics is still in the process of distribution.

The number of respondents to the 2009 survey is significantly higher than past surveys. The over tenfold increase in respondents has many causes. The extension of the survey into more languages contributed greatly to its accessibility, but the largest increase in respondents was a direct result of the decision to include men, who make up over 80% of physicists worldwide. Another part of the increase in responses is likely a result of the greater degree of inclusion of the team leaders in the survey development process. Including the team leaders in the survey development and translation review, in addition to the distribution, helped to foster a good working relationship among the people working on the project and helped to foster a greater sense of connection to the project.

The working group tried to involve the physical societies of various target countries, and at least fourteen team leaders planned to distribute the survey in conjunction with their domestic physical societies, which are popular (and sometimes the only) membership societies for people in physics in each country. Some of the physical societies are simply posting the survey link on their websites or listservs. But two countries, Germany and the United States, are distributing the survey to organized probability samples of their home physical societies. These probability samples may provide some insight into the success of the nonprobability snowball sample, and may in fact provide a model for more countries in future rounds of the survey. Heath (2009) suggests that "random route or quota samples may be quite acceptable alternatives in a country where probability samples could not be implemented effectively, providing they do not involve excessive levels of non-response bias." However, response bias may be extremely difficult to quantify in these circumstances.

The survey distribution has greatly improved from past years, but nonprobability sampling has inherent problems. As with many cross-national surveys that rely on snowball sampling (Katz, 2006), our snowball over-represents developed nations and under-represents scientists from developing nations, who often have greater difficulty contacting their colleagues. Physicists with many professional connections, or working in environments with many physics coworkers, and even physicists working in more urban areas with more dependable sources of electricity are more likely to receive a survey invitation from their colleagues. Contacts are skewed by the characteristics of the distributors, as well as by their degree of participation.

2.4 Analysis

"Differences between countries in response rates and other features of survey design are associated with… substantive outcomes and… ignoring these methodological problems will affect the validity of cross-national comparisons."(Heath et al., 2009)

Past analyses of the Global Survey have included two main components; a quantitative analysis, which both summarized the situation of women overall and contrasted respondents from developed and developing countries, and a qualitative analysis, or group of quotes that illustrate the discrimination and challenges that women face in the field of physics. The wider respondent base of the 2009 survey will push this traditional report format. With the addition of male responses, women's experiences will have a

much deeper context and richer meaning. But there are legitimate concerns about the data, including the nonprobability sample and the issues of cultural equivalence.

One topic covered in the Global Survey that illustrates concerns over cultural equivalence is housework. Housework is described quite vaguely in the questionnaire, *Who is responsible for the majority of the housekeeping in your household?* Batalova & Cohen (2002) demonstrate a gender difference in the perception of housework that is strong enough to warrant a division into routine housework, which is usually done by female members of the household, and occasional housework, which is usually done by male members of the household. This question is problematic in a monolingual environment because it is so strongly influenced by expectations that differ significantly by gender and by subculture, as well as by other complicating factors, such as the size of the household and the presence of children in the house. But it is even more problematic transnationally. Fuwa (2004) found other factors that affect housework, including economic development, female participation in the labor force, gender norms, welfare regimes, and more. Amer et. al (2009) tell the story of an American survey that was distributed in Qatar without first being adapted to Qatari culture. They conclude that "Western-designed questions about family structure and household definitely cannot be imported unmodified into the Qatari context, given a culture in which polygamy and extended families are factors." And yet the importance of this area of study is essential to a thorough understanding of women in physics. Baxter et. al. (1991) suggest that devoting a large proportion of time to housework puts women at a disadvantage in their careers.

Heath et al. (2009) describe the main trouble areas in the analysis of cross-national survey data as data quality (e.g., discrepancies between the population and the sampling frame), nonresponse error (which we cannot approximate) and measurement error (mismatches between reported responses and underlying attitudes). Kohn (1987) cautioned that while it may be appropriate to discuss data in terms of underlying similarities, discussing differences in the data requires expert knowledge of the underlying differences in history and culture. According to Heath et al. (2009), "A major problem in cross-national research is determining whether differences that one finds are real ones or methodological artifacts… The danger lies in the risk of developing substantive theories to explain what is essentially an artifact in the dataset."

Smith (2009) found differing definitions of what constitutes "sensitive topics" and undesirable behavior across cultures. But even item response styles can have a significant influence on the compatibility of responses in multicultural survey efforts (Baumgartner & Steenkamp, May 2001). Some cultures demonstrate an acquiescence bias, with a tendency to choose toned down answers. Some cultures tend to respond using the scalar extremes, and still others make an effort to choose socially desirable responses. Response styles vary not only by culture, but also interpersonally within a given culture and its subcultures. There are a few methods purported to reduce the effect of biases in response style, but within our survey, the best we can do is to try to identify instances of it, correct for it where possible, and frame our findings in ways that openly acknowledge their inherent biases when corrections are not possible.

But cultural differences extend deeper than simply differences in response styles. "If one common variable with the same categories for all countries is desired, then it should be clear that the same categories are highly unlikely to measure exactly the same in different countries" (Braun, 1995). Frijda and Jahoda (1966) explained it well: "If similar activities have different functions in different societies, their parameters cannot be used for

comparative purposes." Our topics of study would have to be thoroughly examined in order to identify this kind of problem. For example, we learned that travel is very important to many scientists in developing countries with limited resources, but in order to report on travel, we would need to understand what travel means in the cultures we wish to describe. In fact, Kohn believes that expert knowledge of all countries involved in a multi-national survey is absolutely necessary (Kohn, December 1987). Kuechler (1998) expressed this doubt: "The current trend of extending cross-national projects to a large number of rather heterogeneous countries is seen as producing data of dubious quality." What does Kuechler's reasoning say about the validity and usefulness of the data we are collecting? Kohn adds to this question: "It is not necessarily true that the more nations included in the analysis, the more we learn. There is usually a trade-off between number of countries and the amount of information obtained." (Kohn, December 1987) Indeed a cursory examination of the quantitative findings of the 2002 and 2005 reports show large differences in even the basic statistics collected, possibly because the 2005 report contained respondents from more countries.

Should we be limiting our transnational surveying to countries where we have a good understanding of the culture and solid cooperation? Would that bias our findings even more? Heath et al. (2009) echoed this concern, but took it further. "Countries for which equivalence of meaning cannot be demonstrated might be dropped from the analysis, although this strategy runs the risk of narrowing the range of countries to more homogenous and possibly westernized ones… Systematic exclusion runs the risk of reducing the theoretical and substantive interest of one's findings. Nevertheless, it is probably always good practice to test the sensitivity of one's conclusions to the exclusion of cases where data quality is expected to be low." Malpass & Poortinga (1986) offer further reprieve: "concepts with functional equivalence are universal in qualitative sense, although not necessary quantitative sense." And Kuechler insists that multi-national research must contain both quantitative and qualitative components (Kuechler, 1998).

In a situation where qualitative data is being used to offer a deeper validity to an analysis of an unfamiliar group, the researcher must be very careful in their approach of the qualitative data. Our goal as researchers is to accurately reflect our findings with as little subjective interpretation as possible. As such, it is important to follow proper qualitative procedures, including multiple coders and a set of agreed upon codes that are only developed after the coders are familiar with the data. It is important to represent the data from the ground up, focusing analysis on the greatest points of concern raised by the respondents, rather than on the research goals of the analysts.

3. Conclusion

Global surveys are inherently difficult and problematic. Special care and attention need to be focused on issues of linguistic and cultural equivalence. A global survey produced in isolation from multinational contacts cannot produce valid results. Smith (2009) advises us to avoid "safari research:" an imperialistic tendency to develop an instrument and "rigidly impose" it on other societies.

Indeed, survey development needs to be approached in an iterative, collaborative, multinational way, with repeated attention toward quality control, in order to result in reliable and valid findings. Multinational, multilingual and multicultural research is especially difficult, but wholly necessary. It is imperfect, but growing. And it is only by looking critically at the strengths and weaknesses of our research that we will be able to

improve it.

Acknowledgements

We would like to thank Kenny Harris for his excellent help on this project, as well as the members of the International Union of Pure and Applied Physics (IUPAP) Working Group on the Status of Women in Physics for their important contributions and foundational support. We would also like to thank the Henry Luce Foundation for its support of the translation, and Karen Kaleuati, Judith Mulvey, Julius Dollison, and Jerry Hobbs for their help with the language review process.

References

Amer, S. R., Ingels, S. J., & Mohammed, A. (Autumn 2009). Validity of Borrowed Questionnaire Items: A Cross-Cultural Perspective. *International Journal of Public Opinion Research* , Vol 21, No 3, pp 368-375.

Batalova, J. A., & Cohen, P. N. (August 2002). Premarital Cohabitation and Housework: Couples in a Cross-National Perspective. *Journal of Marriage and Family* , Vol 62, pp 743-755.

Baumgartner, H., & Steenkamp, J.-B. E. (May 2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research* , Vol. XXXVIII, pp. 143-156.

Baxter, J. (1997). Gender Equality and Participation in Housework: a Cross-National Perspective. *Journal of Comparative Family Studies* , Vol 28, excerpt: http://www.questia.com.

Blasius, J., & Thiessen, V. (2006). Assessing Data Quality and Construct Comparability in Cross-National Surveys. *European Sociological Review* , Volume 22, Number 3, pp 229-242.

Braun. (1995). Some Guidelines for the Measurement of Education in International Projects. *ISSP Field Guide* .

Braun, M. (2003). Communication and Social Cognition. In J. Harkness, F. J. van de Vijver, & P. Mohler, *Cross-cultural Survey Methods* (pp. 57-67). Hoboke, NJ: Wiley.

Braun, M., & Scott, J. (Autumn 2009). Gender-Role Egalitarianism-Is the Trend Reversal Real. *International Journal of Public Opinion Research* , 362-367.

Braun, M., & Scott, J. (1998). Multidimensional Scaling and Equivalence: Is having a job the same as working? *Cross-Cultural Survey Equivalence* , 129-?

Brislin, R. W. (1986). The Wording and Translation of Research Instruments. In W. J. Lonner, & J. W. Berry, *Field Methods in Cross-Cultural Research* (pp. 137-164). Beverly Hills, CA: Sage.

Chaunsheng, C., Lee, S.-y., & Stevenson, H. W. (May 1995). Response Style and Cross-Cultural Comparisons of Rating Scales among East Asian and North American Students. *Psychological Science* , Vol. 6, No. 3, pp 170-175.

Couper, M. P. (2008). *Designing Effective Web Surveys.* NY: Cambridge University Press.

Dillman, D. A. (2007). *Mail and Internet Surveys. The Tailored Design Method.* Hoboken, NJ: Wiley.

Frijda, N., & Jahoda, G. (1966). On the Scope and Methods of Cross-Cultural Research. *International Journal of Psychology* , Vol. 1, No. 2, 109-127.

Fuwa, M. (2004). Macro-level Gender Inequality and the Division of Household Labor in 22 Countries. *American Sociological Review* , Vol. 69, No. 6, pp-751-767.

Harkness, J. AAPOR Short Course (May 13, 2010). Multilingual, Multicultural & Multinational (3M) Survey Design Methods.

Harkness, J. A., & Schoua-Glusberg, A. (January 1998). Questionnaires in Translation. *ZUMA-Nachrichten Spezial* , 87-126.

Harkness, J., van de Vijver, F. J., & Mohler, P. P. (2003). *Cross-Cultural Survey Methods.* New York: Wiley.

Heath, A., Fisher, S., & Smith, S. (June 2005). The Globalization of Public Opinion Research. *Annual Review of Political Science* , Vol. 8, pp 297-333.

Heath, A., Martin, J., & Spreckelsen, T. (Autumn 2009). Cross-National Comparability of Survey Attitude Measures. *International Journal of Public Opinion Research* , 293-315.

Heckathorn, D. (2006). *Respondent Driven Sampling*. Retrieved 2009, from http://www.respondentdrivensampling.org/

Ivie, R., & Guo, S. (April 2006). *Women Physicists Speak Again.* American Institute of Physics, AIP Pub Number R-441.

Ivie, R., Czujko, R., & Stowe, K. (2002). Women Physicists Speak. The 2001 International Study of Women in Physics. *Women in Physics: The IUPAP International Conference on Women in Physics* (pp. 49-74). Melville, NY: American Institute of Physics.

Johnson, T. P. (January 1998). Approaches to Equivalence in Cross-Cultural and Cross-National Survey Research. *ZUMA-Nachrichten Spezial* , 1-3.

Jowell, R. (1998). How Comparative is Comparative Research. *American Behavioral Scientist* , Vol. 42, No. 2, pp. 168-177.

Kalleberg, A. L., & Rosenfeld, R. A. (May 1990). Work in the Family and in the labor Market: A Cross-national, Reciprocal Analysis. *Journal of Marriage and the Family* ,

Vol 52, pp 331-346.

Katz, H. (2006, November 9). Global Surveys or multi-national surveys? On sampling for global surveys. *Thoughts for the Globalization and Social Science Data Workshop UCSB* .

King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (February 2004). Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. *American Political Science Review* , Vol. 98, No. 1, pp 191-207.

Kleiner, B., Pan, Y., & Bouic, J. (2009). The Impact of Instructions on Survey Translation: An Experimental Study. *Survey Research Methods* , Vol.3,No.3, pp. 113-122.

Kohn, M. J. (December 1987). Cross-National Research as an Analytic Strategy. *American Sociological Association 1987 Presidential Address* , Vol. 52, pp 713-731.

Kuechler, M. (1998). The Survey Method. *American Behavioral Scientist* , Vol. 42, No. 2, pp 178-200.

Lynn, P. (2003). Developing quality standards for cross-national survey research: five approaches. *International Journal of Social Research Methodology* , Vol. 6, No. 4, pp 323-336.

Malpass, R. S., & Poortinga, Y. H. (1986). Strategies for Design and Analysis. In W. J. Lonner, & J. W. Berry, *Field Methods in Cross Cultural Research* (pp. 47-83). Beverly Hills, CA: Sage.

Malpass, R. S., & Poortinya, Y. H. (1986). Making Inferences from Cross-Cultural Data. In W. J. Lonner, & J. W. Berry, *Field Methods in Cross-Cultural Research* (pp. 17-46). Beverly Hills, CA: Sage.

Medina, T. R., Smith, S. N., & Long, J. S. (Autumn 2009). Measurement Models Matter:Implicit Assumptions and Cross-National Research. *International Journal of Public Opinion Research* , 333-361.

O'Muircheartaigh, C. (April 11, 2008). Survey Design a la mode. Survey Research in the 21st Century. *JPSM Distinguished Lecture* .

Park, A., & Jowell, R. (1997). Consistencies and Differences in a Cross-National Survey. *SCPR* .

Pescolido, B. A., & Olafsdottir, S. (June 25-29, 2008). The Logistics of Survey Implementation in a Comparitive Study of Mental Illness: Issues and Resolutions in Translation Across Cultural Boundaries. *Prepared for Session 33 (Health and Mental Health), International Conference on Survey Methods in Multinational, Multiregional, and Multicultural Contexts.* Berlin.

Potaka, L., & Cochrane, S. (2004). Developing Bilingual Questionnaires: Experiences from New Zealand in the Development of the 2001 Maori Language Survey. *Journal of*

Official Statistics , Vol. 20, No. 2, pp. 289-300.

(1970). Chapter 2: Research Designs. In A. Przeworski, & H. Teune, *The Logic of Comparaitive Social Inquiry* (pp. 31-46). New York: Wiley.

Przeworski, A., & Teune, H. (Winter 1966-67). Equivalence in Cross-national Research. *Public Opinion Quarterly* , Vol. 30, No. 4, pp 551-568.

Schaeffer, Nora Cate. JPSM Short Course (April 12-13, 2007). Questions for Standardized Measurement in Surveys

Schoua-Glusberg, A. (2004, July 7). Translation of Survey Instruments. *Presentation to DC-AAPOR* . Washington DC.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation. John M. MacEachran memorial lecture series, Vol. 1995.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers. xi, 112 pp.

Scott, J., Alwin, D. F., & Braun, M. (1996). Generational Changes in Gender-Role Attitudes in Britain in a Cross-National Perspective. *Sociology* , Vol. 30, No. 3, pp 471-492.

Smith, T. W. (2009). Developing and evaluating cross-national survey instruments. In J. M. Presser, *Methods for Testing and Evaluating Survey Questionnaires* (pp. 69-91). Hoboken, NJ: John Wiley & Sons, Inc.

Smith, T. W. (Autumn 2009). Editorial. *International Journal of Public Opinion Research* , Volume 21, Number 3, pp. 267-270.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response.* New York: Cambridge University Press.

Trea, J., & Widmer, E. D. (June 2000). Married Women's Employment over the Life Course: Attitudes in Cross-National Perspective. *Social Forces* , Vol. 78, No. 4, pp 1409-1436.

van de Vijver, P. J., & Poortinga, Y. H. (December 1982). Cross-cultural Generalization and Universality. *Journal of Cross-Cultural Psychology* , Vol. 13, No. 4, pp 387-408.

Verma, V. Comparability in International Survey Statistics. *Keynote Address, International Conference on Improving Surveys, Copenhagen, 25-28 August 2002.*

Vitray, K. (2007). *The New ASTM Translation Guide.* Austin, TX: McElroy Translation.

Watkins, D., & Cheung, S. (September 1995). Culture, gender and response bias: An analysis of responses to the Self-Description Questionnaire. *Journal of Cross-Cultural Psychology* , Vol. 26, No. 5, pp 490-504.

Yu, J., Coons, S., Draugalis, J., Ren, X., & Hays, R. (June 2003). Equivalence of Chinese and US-English Versions of the SF-36 Health Survey. *Quality of Life Research* , Vol. 12, No. 2, pp 449-457.