Opportunities and challenges in the expansion of cross-national survey research

Shawna N. Smith a, Stephen D. Fisher b & Anthony Heath c

a Department of Sociology, Indiana University, Bloomington, IN, USA
b Department of Sociology, University of Oxford, UK
c Department of Sociology, University of Oxford and Institute for Social Change, University of Manchester, UK

Available online: 25 Oct 2011

To cite this article: Shawna N. Smith, Stephen D. Fisher & Anthony Heath (2011): Opportunities and challenges in the expansion of cross-national survey research, International Journal of Social Research Methodology, 14:6, 485-502

To link to this article: http://dx.doi.org/10.1080/13645579.2011.611386

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Opportunities and challenges in the expansion of cross-national survey research

Shawna N. Smith*, Stephen D. Fisher and Anthony Heath

aDepartment of Sociology, Indiana University, Bloomington, IN, USA; bDepartment of Sociology, University of Oxford, UK; cDepartment of Sociology, University of Oxford and Institute for Social Change, University of Manchester, UK

(Received 15 May 2011; final version received 19 July 2011)

Past decades have witnessed significant expansion in the reach of public opinion surveys. While the need for comparable data from increasingly heterogeneous countries has led some to infer that survey practices must be centralized and identical, the explicit inclusion of variation between countries may well improve the enterprise of cross-national research. This paper examines some of the challenges endemic to the continued global spread of cross-national research, both at the level of cross-national survey bodies and the analyst using second-hand data. With regard to the former, we examine recent changes in the ‘product’ of comparative survey research, and the increasing recognition that improvements in both survey quality and equivalence of meaning may come from localized understandings and specifications. We also explore how individual analysts might provide important substantive and methodological insights into the broader enterprise of comparative work, further advancing decolonized methodologies by devolving methodological innovation from survey developers located in prosperous countries to analysts with second-hand data spread across a broader geography.

Keywords: survey methodology; cross-national research; comparative research

Introduction

If Durkheim ([1895]1964, p. 139) is correct, and comparative sociology is ‘sociology itself,’ then cross-national and cross-cultural surveys increasingly provide the lifeblood for social science disciplines. Of course, to the extent that cross-national opinion research attempts to understand social phenomena by systematically comparing across diverse countries, the nobility of the goals are exceeded only by the significance of the challenges.

For survey data from different countries to be meaningfully compared they must be sufficiently comparable. Unfortunately, this requirement has at times been thought to imply that identical surveys must be uniformly fielded – using identical methods – in all countries. Given that most cross-national survey programs have started in the West, the further implication has been that these questions and methods being uniformly applied represent a new form of ‘colonization’ – the unthinking exportation of methodologies developed in the West to countries with vastly different cultures, histories, and purviews.

*Corresponding author. sns3@indiana.edu
As we noted in Heath, Fisher, and Smith (2005, p. 298), the globalization of survey research ‘has led to the dominance of a particular intellectual framework and set of assumptions that may not be quite so appropriate outside their original homes’. Moreover, in some cases variation between countries – for example, in using different sampling frames or modes of data collection – may well improve the quality of cross-national data.

The risk of relying on a one-way valve of survey methodology, from the West to the non-West, is the potential for missing out on the benefit of the methods, insights, and intellectual frameworks that have emerged in the non-Western world. While researchers are often understandably keen to see how individuals in other countries respond to a fascinating question devised in their home country or region, a boilerplate exportation of questions or concepts, without a view as to how that question might be interpreted or understood differently, gives the potential for non-Western imports and insights short shrift.

This article examines some of the opportunities and challenges endemic to cross-national research – both at the level of the major cross-national survey research bodies and the analyst using second-hand data. With regard to the former, we examine recent changes in the ‘product’ of comparative survey research. As we discussed in 2005, the globalization of public opinion research does not necessarily imply the spread of an identical product. Both within and between comparative surveys, important differences exist with regard to issues of measurement (e.g., coverage and sampling) and equivalence (e.g., translation and meaning). Increasingly, programs are being adopted to incorporate understandings and insights of non-Western countries explicitly into data collection protocol – ‘decolonizing’ certain survey practices.

With a few exceptions, analysts have been neglected in discussions of cross-national research quality. The penultimate section suggests several strategies analysts should consider in ensuring reliable and valid comparative research findings, especially as they find themselves using data-sets containing increasingly diverse countries and cultures. We also highlight how individual analysts can provide important substantive and methodological insights for the leaders of large comparative survey research projects – further advancing decolonized methodologies by devolving methodological innovation from survey developers located in prosperous countries to analysts spread across a broader geography. Further methodological dispersion may, in fact, rely on individual analysts.

Cross-national surveys and methodological standards

Quality issues

As we noted in 2005, ‘[A]lthough public opinion research has clearly been affected by the pursuits of globalization, only very recently could it be considered a global phenomenon’ (311). In the past six years, the global reach of surveys has continued to grow, both through the expansion of pre-existing surveys and the establishment of new cross-national programs.

Much of this growth has centered on bettering the coverage and representation of the non-Western world in cross-national surveys. Several of the newest survey series, in fact, cover exclusively non-Western countries – for example, the Arab Barometer and South Asian Barometer, both established in 2005, and the East Asia Social Survey. Amongst the more ‘global’ surveys – both academic and non-academic – waves now regularly include countries from Africa, the Middle East,
Latin America, and Asia. Wave 5 of the World Values Survey (WVS) included its largest proportion of non-Western countries to-date, and the global Gallup and Pew survey series include increasingly diverse representation of non-Western countries.

This continued geographical spread has necessitated continual adaptations in methodology. Heath et al. (2005) discussed some of the methodological challenges inherent in working with ‘globalized’ cross-national data. Other, more recent, volumes have further investigated these challenges – in particular, de Leeuw, Hox, and Dillman (2008), Harkness et al. (2010), Haller, Jowell, and Smith (2009), and Jowell, Roberts, Fitzgerald, and Eva (2007). Further, these volumes have all highlighted the importance of pursuing cross-national research, particularly as more numerous and diverse countries join in the mix. In recognition of these complications, several noted scholars have called for new and innovative understandings of ‘quality’ in cross-national research, to provide research programs with an incentive to look beyond long-surveyed countries with well-established norms (e.g. Harkness, 1999; Lynn, 2003).

Following Heath et al. (2005), this section considers developments in survey practice, particularly as they have evolved to accommodate larger and more diverse sets of countries, first considering issues of survey quality and then issues of equivalence.

A particular concern for cross-national researchers is the extent to which errors in survey quality may vary between countries and lead to cross-national differences that are purely methodological artifacts (Heath, Martin, & Spreckelsen, 2009). Groves (1987) identifies the four most commonly discussed kinds of survey error: (1) coverage, or a failure to ensure that all portions of a population were adequately represented; (2) nonresponse, or differences in response rates between countries; (3) sampling, or errors that result from a failure to gather a large enough sample (within-country) or from differences in sampling strategies across countries; and (4) measurement, or errors resultant from the influence of the interviewer, weakness of the survey questions, failures of the respondent to give appropriate answers to the questions, and effects of the mode of data collection on survey answers.

When it comes to discussions of survey quality – even as surveys expand at increasing rates – interest is often focused on errors in practice rather than conceptualization. In other words, scholars of cross-national and cross-cultural work are generally in agreement that, for example, full coverage is always better than partial, and that random samples are preferable to quota samples. Nonetheless, survey bodies themselves are responsible for defining their standards and, in turn, establishing which aspects of survey quality are most important.

Of the major cross-national surveys, the European Social Survey (ESS) makes the strongest effort to minimize these different sources of error. Even as it has expanded, the ESS has insisted on participating countries adhering to a rigorous set of standards. With regards to coverage, the ESS sets its target population as ‘persons 15 years or older who are resident within private households, regardless of nationality and citizenship or language’ (ESS, 2010, p. 1). The ESS has set a target response rate of 70%, and provides extensive guidance materials for coordinators and interviewers on, e.g. refusal avoidance and incentivization (Koch, Fitzgerald, Stoop, & Widdop, 2010). Participating countries must also achieve a minimum effective sample size of either 1500 or 800, depending on population, and non-probability samples – including quota samples, substitution, and random-route
sampling – are forbidden. Only face-to-face interviews are used (although supplementary questionnaires may be completed by the respondent with an interviewer present) and survey completion time is kept short to improve temporal comparability.

The ESS has had much, if not perfect, success in attaining their methodological goals. Survey documentation indicates that participating countries have adhered to coverage requirements, met minimum effective sample sizes, and followed full probability sampling requirements. Round 4 ESS reports indicate that although a majority of countries failed to reach the 70% response rate benchmark, the number approaching or reaching this benchmark is improving (Matsuo, Billiet, Loosveldt, & Malnar, 2010).

Of course, the ESS is more bounded geographically than other ‘global’ survey programs, which have to consider trade-offs between quality and expansion on a broader scale. Table 1 summarizes the approaches taken by the ESS and three major ‘global’ survey programs – namely, the International Social Survey Programme (ISSP), the WVS and the Globalbarometer surveys – in defining ‘quality’ standards. As this table shows, the global surveys, while not as rigorous as the ESS, pay heed to similar goals of full coverage, minimum sample sizes, and full probability samples. Much like the ESS, the WVS and ISSP have instituted minimum sample sizes, and the ISSP and various Globalbarometer surveys have also instituted programs of pre-testing, largely centered on ensuring adequate translation. All of the survey programs have privileged face-to-face interviewing over alternatives. In fact, face-to-face administration is one of the few requirements of the decentralized Globalbarometer survey project – most other practices are determined by the regional barometers themselves – and is also listed as a central tenet of the WVS constitution. Perhaps even more telling, although the ISSP designs its surveys for self-administration, nearly 2/3 of participating countries in 2009 administered surveys face-to-face. Such reports seem to indicate that face-to-face administration is seen as the most reliable method across the widest range of countries.

Table 1 also makes clear some divergence in standards, particularly with regard to coverage and sampling requirements. Much like the ESS, the WVS has also established a specific target population on the basis of age, citizenship, and residence. While the ISSP does not specify an overall target population, it does report on the target for each country, again in terms of age, citizenship, and residence requirements. Further, while the ESS requires strict probability samples, the ISSP and WVS do not. Nonetheless, there is some indication of a decline in the number of countries in these surveys using non-probability samples – in the 1998 ISSP, for example, 12 of 24 countries used some form of non-probability sampling (Harkness, Langfeldt and Scholz, 2000), while in 2006 such sampling was used in only 2 countries of 33 (Scholz, Faab, Harkness and Heller, 2008).

Of course, probability samples require not only academic resources and infrastructure, but – in some cases – a change in tradition. Harkness (1999) points out that quota sampling has long been the norm in France and, indeed, France is one of only five countries in Round 4 of the WVS to use quota samples (Medrano, 2009).

Interestingly, a survey-wide commitment to probability sampling has given rise to one of the most localized aspects of cross-national surveys – variation in sample design. Kish (1994) argues that it is both necessary and preferable to establish the sample, including stratification design and the establishment of the sampling frame, at a more ‘local’ level because differences in resources between countries lead to
Table 1. Established methodological standards amongst major cross-national academic surveys, by survey organization.

<table>
<thead>
<tr>
<th>Survey Organization</th>
<th>Coordination?</th>
<th>Pre-testing?</th>
<th>Minimum sample size?</th>
<th>Mode of administration</th>
<th>Coverage requirements</th>
<th>Sampling requirements</th>
<th>Response rate target?</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESS</td>
<td>Strong central coordination and design, participating nations fund fieldwork through scientific bodies.</td>
<td>YES: Quota sample of 50 in each country, debriefs and cognitive interviewing.</td>
<td>YES: Effective sample of 1500 (or 800 in small countries) after design effects.</td>
<td>Face-to-face only, self-completion allowed for supplementary questionnaire in interviewer presence.</td>
<td>Full coverage of residential population aged 15+ (no upper age limit), weights required.</td>
<td>Full probability only, no ‘random route’, oversampling may be allowed.</td>
<td>YES: 70%, four personal visits requested.</td>
</tr>
<tr>
<td>ISSP</td>
<td>Centrally coordinated, participating nations self-fund.</td>
<td>YES: In 2007, 15 of 29 non-English speaking countries pre-tested in at least one other language.</td>
<td>YES: 1000 with aim of 1400.</td>
<td>Designed for self-completion, in 2007, 23 of 34 were exclusively face-to-face.</td>
<td>Documentation only. Differences in citizenship requirements, age restrictions, and accommodation present.</td>
<td>Documentation only. Non-probability (quota sampling &amp;/or substitution) allowed but must be reported.</td>
<td>NO: Reporting required following AAPOR/WAPOR standards. Required visits in 2007 ranged from 0 to 10.</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Coordination?</th>
<th>Pre-testing?</th>
<th>Minimum sample size?</th>
<th>Mode of administration</th>
<th>Coverage requirements</th>
<th>Sampling requirements</th>
<th>Response rate target?</th>
</tr>
</thead>
<tbody>
<tr>
<td>WVS(^c)</td>
<td>Central governing body, participating nations self-funded.</td>
<td>NO: None discussed or documented.</td>
<td>YES: 1000.</td>
<td>Constitution specifies face-to-face, all but three countries report face-to-face for Round 5.</td>
<td>Individuals 18 and over, regional (e.g. urban-only) surveys not included in aggregate data files.</td>
<td>Documentation only, requirement of ‘as close as possible’ to full probability sampling. Some quota in Round 5.</td>
</tr>
</tbody>
</table>

\(^a\)ESS (2009, 2010).
\(^b\)Scholz & Heller (2009).
\(^c\)Mission Statement and Goals of the World Values Survey Association (n.d.).
different optimal sampling frames or sampling designs. Thus, while a postcode address file might provide a thorough, easily accessible sampling frame in a more established Western country, sampling from a postcode address file in a less well-off country could lead to systematic under-representation of certain social classes. In line with this understanding, then, the ESS (2010, p. 1) recommends, ‘the best random practice used in each participating country. The choice of a specific design depends on the availability of frames, experience, and of course also the costs in different countries’. The ISSP and WVS also allow for countries to determine optimal sampling strategies, but require them to document these strategies fully. Of course, choosing an ‘optimal’ sampling frame for a country presupposes the availability of any suitable sampling frame – something not all countries currently have. Even within Europe, for example, there is no publicly accessible nationwide sampling frame available in Bosnia and Herzegovina, which is troubling given that such countries may well provide some of the sternest challenges to Western theories.

The establishment of locally-optimal sampling protocols also highlights the emerging importance of documentation for cross-national surveys. Both content of and access to cross-national survey documentation have improved quickly over time. As Mohler, Pennell, and Hubbard (2008, p. 403) note:

In just a few decades, paper documents describing the contents of rectangular data files (i.e. codebooks) have been replaced with online access to documents that both describe and facilitate analysis of complex hierarchical and/or relational databases. Despite these advances, examples of complete or even adequate survey documentation remain surprisingly rare.

While some surveys, like the ESS, document procedures centrally, most surveys rely on individual countries to submit reports on data collection. By keeping documentation at a more localized level, countries are provided with an opportunity to describe their sampling frames, sample designs, decisions on populations and coverage, etc. through a more local understanding. However, sometimes this lack of centralization means that certain countries lack documentation, either completely or in part. For example, country-specific documentation of Wave 5 of the WVS has significant inconsistencies in whether or how things like coverage or sampling procedures are reported (Medrano, 2009); similarly, the 2007 ISSP Study Monitoring Report indicates that only 20 of 34 countries provided specific methodology reports (Scholz & Heller, 2009). Despite these missing reports, the ISSP provides a good model for cross-national survey documentation, with country reports provided alongside an ISSP-sponsored ‘study monitoring report.’ Together these reports provide accurate yet concise summaries of the locally appropriate strategies and procedures followed by each surveyed country.

*Equivalence of meaning*

In 2005, we noted that while cross-national surveys generally allow for some dissimilarity between countries in survey design and implementation, they also privilege the standardization of questionnaire content. However, as the paradigm of ‘functional equivalence’ has come to the fore, new questions have arisen about
which survey aspects should be standardized and which should not be – or how processes of ‘standardization’ should be codified or considered.

Issues relating to translation have long been central to concerns about equivalence, as differences in conceptual understanding, word meaning or even dialect can affect fundamental item comprehension. The earliest solution to translation issues – iterative translation and ‘back translation’ – quickly drew criticism, with numerous scholars noting that such procedures overlooked larger concerns of linguistic equivalence (e.g. Harkness, 1999; Harkness & Schoua-Glusberg, 1998; Warwick & Osherson, 1973).

Few ‘solutions’ for ensuring equivalence across a set of diverse countries have been suggested. With regard to translation issues, Harkness (2003, 2007) recommends the five-step ‘TRAPD’ procedure, involving: (1) translation, (2) review, (3) adjudication, (4) pre-testing, and (5) documentation. This process is designed to involve multiple skilled translators and survey specialists within each country working to arrive at an optimal translation. A recent evaluation of this method across a variety of types of cross-national and cross-cultural studies suggests that it does aid the identification of translation problems (Willis et al., 2010).

Cognitive interviewing provides one of the most promising roads for exploring – and ideally ensuring – equivalence across countries. The cognitive interviewing process provides individuals with an opportunity to discuss in their own words how they interpret specific questions or responses, as well as explain why they responded in a particular way. More broadly though, cognitive interviewing allows survey researchers to explore whether the intellectual framework of the survey, or the implicit assumptions of the question or concept, are in line with the intellectual framework or assumptions of the respondents. As surveys traverse into increasingly diverse territories, such knowledge is perhaps the only way to ensure that survey questions are reliable and valid – not simply for one population of interest, but for all populations of interest.

Although earlier forays into cognitive interviewing proved less than successful (Miller, 2007), the ESS has introduced a program of cognitive pre-testing examining translations and meanings across six countries and seven languages. Reports on this initial round include examples of the identification and correction of various types of errors across surveys, thus diminishing the need for post hoc correction (Fitzgerald, Widdop, Grey, & Collins, 2009). Certainly, the adoption of such techniques by survey programs with a broader geographical purview would provide interesting insights into survey effectiveness and differences between diverse countries of interest, further facilitating the decolonization of centralized survey programs.

The ESS has also led the charge in exploring ways of optimizing post hoc equivalence, specifically via ‘Multi-Trait Multi-Method’ (MTMM) studies, where respondents are asked either multiple versions of the same question or are randomly assigned to one version of a question (typically referred to as Split Ballot MTMM, Saris & Gallhofer, 2007). The resulting data allow researchers to model differences in error variances that might occur between countries, languages or cultures, and thus to correct for these differences. In practice, corrections based on MTMM have been important enough to affect substantive country-level differences (Saris & Gallhofer, 2007, n.d.); again however, exploration of this method by a survey not confined to Europe may be necessary before we appreciate its full methodological value.
Concerns of ‘differential item functioning’ (DIF), where subjective response categories are used differently in different countries due to, e.g. cultural emphases on different personality traits (e.g. Uskul, Oysterman, & Schwarz, 2010), add a further level of complication to ensuring equivalence of meaning, in that both questions and potential responses are susceptible to bias (Schwarz, Oysterman, & Peytcheva, 2010; Yang, Harkness, Chin, & Villar, 2010).

In 2005, we discussed King, Murray, Salomon, and Tandon’s (2004) use of ‘anchoring vignettes’ as a solution for DIF by providing a means to estimate how response options on a question are used differently in different countries. While the anchoring vignettes idea has grown in popularity and been included in several systematic cross-national surveys – including the World Health Organization’s World Health Survey – such a method has yet to be adopted by any of the major cross-national public opinion surveys.

The introduction of advanced translation techniques, cognitive interviewing and vignettes all emphasize the importance of local variation in meaning in ensuring methodological equivalence; however, at present most of these methods have not been utilized by the larger cross-national series. Applications of TRAPD, MTMM, and cognitive interviewing have been limited, examining only a subset of countries and/or indicators; there remains much room for both diffusion and innovation. In addition, in 2005 we noted that, ‘a program of methodological work to determine which measuring instruments are particularly prone to problems of “equivalence of meaning” appears overdue’ (Heath et al., 2005, p. 324). We reiterate this call here. Furthering our understanding of which questions might travel best, particularly across more heterogeneous cultures and contexts, is crucial to a methodologically responsible program of globalized public opinion research.

Devolving decolonization: the role of the cross-national analyst

In his classic piece on credibility and comparative research, Verba (1971, p. 327) remarked,

\[\ldots \text{The very process of data analysis can be considered a ‘bootstrap’ operation. In the process of examining data from a substantive point of view, one can also increase the credibility of the results. Such an approach to increasing credibility of comparison has a particular advantage from the point of view of the scholar conducting secondary analysis of survey data. Where the scholar has no control over the data-gathering process, he can still exercise some ‘data quality control’ over the results.}\]

In spite of this early consideration, more recent discussions of cross-national research challenges have largely ignored the potential for secondary-data analysts to advance methodology. Such an oversight is interesting for a number of reasons. For one, the sheer volume of secondary analysts could lighten the load on survey programs, at least with regard to investigating certain post hoc phenomena. Second, and perhaps more importantly, the geographical dispersion of secondary-data analysts provide survey bodies with built-in country-specific experts. The most recent ISSP bibliography, for example, includes hundreds of papers written about non-Western countries, frequently by native scholars publishing in native languages (Smith, 2010). Rather than viewing the reliability and validity of cross-national data as purely the job of survey teams, secondary-data analysts...
hold the potential to both diminish the risk of publishing erroneous findings and provide helpful, substantively informed and culturally specific feedback for cross-national survey organizations.

This section outlines some steps analysts can take to both improve the validity of their analyses and help ensure findings that are not methodological artifacts. By being both aware of and resourceful in addressing potential shortcomings, analysts may spearhead much of the methodological innovation coming from cross-national work. In this sense, we speak of the potential for a devolved decolonization of cross-national survey data.

**A renewed call for meta-analyses and other a-priori considerations**

**The need for meta-analyses**

In our 2005 piece, we drew from evidence-based medicine to propose a system of weighting survey research, with the highest weight going to meta-analyses of systematic samples with full documentation (Heath et al., 2005, p. 328). While some such meta-analyses have been published, for example, those on electoral turnout by Geys (2006) and Blais & Aarts (2006), there is scope for many more. So we reiterate this call. Much as astronomers use data from multiple detectors to assess the degree to which the detection of a new planet or object is separate from instrument-based effects, scholars with an interest in detecting cross-national differences may need to become accustomed to using multiple survey instruments to identify more robust findings.

Similarly, as cross-national programs extend to more and more waves, analysts should consider cross-national findings across time. Barring exogenous political or organizational shocks, the ability to replicate results or findings of older, more established data-sets provides some face validity for newer or less established instruments or studies – something that may be particularly important as more region-driven surveys, like the regional Barometers, are developed to cover non-Western countries. Similarly, if a prior result is not replicable, or relationships appear to change significantly over short periods of time, this may be a reason to investigate differences in data composition or methodological procedure, or explore alternative data sources. Again, such results might be especially helpful in non-Western countries with identifying problematic survey items or exploring correlations between response patterns and, for example, sampling strategies.

Related to meta-analysis of previously published research is the analysis of regression output conducted on a survey-by-survey basis. The two-step approach (Jusko & Shively, 2005) is an alternative to the multilevel analysis of a pooled data-set comprising surveys from different countries (or perhaps also different times). The first step is to conduct regression analyses on each survey separately and the second step is to analyze the resulting coefficients of key explanatory variables from the first step. One advantage of this approach over multilevel modeling is that it provides flexibility in the model specification across surveys to allow for sensitivity to local context (e.g. Duch, May, & Armstrong, 2010; Duch & Stevenson, 2006) – again allowing for differences between Western and non-Western countries to become both apparent and explorable.
Data selection

Analysts use many substantive criteria in selecting their data source, not least requirements for particular survey questions and country coverage. Beyond these there may be scope for data selection on the grounds of data quality, for example, giving preference to those surveys that use full probability samples or have more complete documentation. For those data that are selected (perhaps all available) it is important to understand the key methodological features of the surveys, including coverage, response rates, design effects, and potential issues of translation or equivalence.

Of course, secondary-data analysts are not restricted to one survey program – or in fact any cross-national survey program – to do cross-national work. A great deal of high-quality comparative work has been done using country or region-specific surveys (e.g. Diamond & Plattner, 2008; Duch et al., 2010; Heath & Cheung, 2007). As the menu of countries and substantive areas covered by different survey programs grows, analysts should consider optimizing individual-country quality by comparing across different surveys. This approach, of course, comes with its own potential shortcomings – for example, the risk of measurement error induced by differences in question wording, question order, or interviewer procedure – but such shortcomings do not necessarily outweigh other potential advantages.

Further, single-country studies, which will typically have been designed by local teams, can provide essential checks on the robustness of findings for that country emerging from a cross-national survey. This can be especially useful where a particular country appears to be an ‘outlier’ in the cross-national analysis. In this way, the local teams can – by proxy – hold the central cross-national team to account, by scrutinizing or challenging unexpected results and interpretations about their own country.

Finally, scholars may want to consider making use of local experts in ruling out or identifying methodological artifacts. As other scholars (Harkness et. al., 2010; Heath et al., 2009; Verba, 1971) have noted, systematic cross-country differences in methodology can be thought of as alternative hypotheses to explain significant differences. While the analyst may not be able to rule these alternate hypotheses out entirely, local experts, with their more detailed knowledge of their country and its survey methods, will be better placed to identify likely methodological artifacts.

Project design

As is often discussed (e.g. Heath et al., 2009; King et al., 2004; Medina, Smith, & Long, 2009; Meredith, 1993; Steenkamp & Baumgartner, 1998; Verba, 1971), relationships between variables may be less susceptible to certain types of measurement error than mean-level comparisons. For example, the contention that support for democracy in Algeria is higher than support for democracy in Yemen could be confounded by a number of issues of survey quality, including differences in coverage, sampling, or measurement error by country. Differences in the patterns of association between variables, however, are perhaps not so affected by these types of error. As such, analysts exploring countries where differences in method could be an issue may wish to privilege differences in relationships between variables across countries or locales over mere differences in mean-level responses.
Errors of non-observation and nonresponse bias

Errors of non-observation, resulting from country-level differences in response rate, coverage, sample design, and/or mode of administration, present one of the most serious risks of artifactual findings. Although surveys routinely report response rates, as Heath et al. (2009) note, it is nonresponse bias—not response rates, per se—that should be of concern to researchers, as bias is not necessarily related to response rates, and countries with high response rates may be just as prone to bias (Groves, 2006).

Billiet (2007) suggests several ways analysts might check for response bias. One of the simplest is comparing distributions from sample data to available population data, such as Census data, for demographic variables like age and education. Heath et al. (2009) compared age profiles in 2003 ISSP data with country-specific Census data and found that the countries using self-completion surveys tended to have the largest total deviations. Such checks require little extra effort on the part of the analyst—indeed, some surveys, notably the ESS and the WVS, include population statistics as part of their documentation—and can provide the analyst with a great deal of information about potential sources of bias.

The application of post-stratification weights is a common way to adjust for known discrepancies between the distributions of certain variables in the sample and the corresponding distributions in the target population. While many cross-national surveys provide post-stratification weights they are not a panacea for nonresponse bias. In particular, as weights can only be derived from observed variables, and sources of bias may well be resultant of unobserved selection patterns, weights may be more problematic in non-Western countries than in countries where relationships between population parameters and observed variables are better established. Any weighting strategy necessarily involves certain assumptions, and analysts need to be aware that assumptions that seem plausible in a Western context may not be so plausible elsewhere. As such, including weights may do little to mitigate overall bias in a model (Heath et al., 2009).

Choosing variables and using scales

Lack of equivalence of survey questions across different contexts remains one of the main sources of potential bias. Even here, there is scope for analysts to explore and diagnose—if not necessarily correct for—problems of equivalence.

Differences in item saliency and nonresponse

As social scientists have long noted, the development of meaningful public opinion depends on exposure to political or economic issues in such a way as to allow the development of individual opinions, attitudes or beliefs about a particular topic. As issues in public debate vary between countries, often systematically by, for example, level of development or economic system, so the scope of measurable social attitudes will also vary.

As an example, the 2002 Pew Global Attitudes survey included the question: ‘There has been a lot of talk about globalization these days. Do you think that globalization is a very good thing, somewhat good, somewhat bad or a very bad thing?’ The percentage of those who did not answer or said ‘Don’t Know’ ranged from 4% in France to 70% in Uzbekistan, with some unsurprising evidence of systematically higher levels of nonresponse in lesser developed countries (Pew Global Attitudes
‘Views of a Changing World,’ 2003). Variation in item-response rates could provide an important insight into the extent to which respondents understand survey questions and, as such, an important precursor to analyzing non-missing responses. Indeed, it is quite likely that many survey respondents in non-Western countries use the ‘Don’t Know’ option to express the fact that they do not know what the question means as opposed to the normal Western assumption that respondents understand the question but do not have a view on the topic.

Scales
The aim of scaling is to both reduce measurement error and cover different aspects of an abstract concept by combining different survey items through some form of measurement modeling. Exploring measurement modeling in the cross-national context holds significant potential for secondary analysts, largely because such models can only be done post-data collection. Just as there is significant room for advancement and improvement amongst survey bodies in exploring and ensuring equivalence of concepts and indicators across countries, there remains much latitude for analysts to explore similar issues in their own work.

Recently analysts have begun to problematize the use of scaling in cross-national work, either by exploring differences across different scaling techniques or by questioning assumptions of measurement models, such as factorial invariance, or the idea that one measurement model works equally well for all countries or all groups. Multiple-group confirmatory factor analysis has been used effectively to show that models of a given concept may differ significantly between countries, regions or groups (e.g. Agarwala and Lynch, 2006; Brown, 2006; Perreira, Deeb-Sosa, Mullan Harris & Bollen, 2005; Vandenberg & Lance, 2000), thus casting doubt on whether a singular measurement model is acceptable for use in secondary analysis. Medina et al. (2009) show that substantive findings of difference between pairs of countries can differ depending on the scaling approach used. Similarly, Braun and Johnson (2010), in an overview of methods for detecting difference, show that the adoption of different measurement model methods can reveal different aspects or problems within the same data.

Given these differences, Medina et al. (2009) call on analysts to better justify the decisions they make regarding scaling, and ideally to theorize explicitly about similarity or dissimilarity in scales across countries or regions. In some cases, the analyst may believe that there is enough evidence to justify their use of a singular measurement model on pooled data – they may, for example, find a model or core group of indicators that is measurement-invariant across the countries of interest. Alternatively, the analyst may opt to use different measurement models in different countries, referred to as ‘locally conditioned’ models, representing the same concept. In these cases, the analyst should be prepared to support this decision using both country-specific theory and data. Once again, local experts who are familiar with the specific country and its previous research will be crucial in the formulation of these locally conditioned models.

Modeling variation
In addition to identifying and reflecting on methodological differences between surveys, analysts may also choose to estimate their effects statistically. For example,
using ISSP data, Heath et al. (2009) found that factor scores for particular survey items in a scale measuring national identity were correlated with the extent of unrepresentativeness of a survey with respect to the distribution of respondent age, reported vote bias, nonresponse rate, and mode of administration.

In addition to—or perhaps in lieu of—explicitly modeling the effects of differences in nonresponse or survey design, Heath et al. (2009) also suggest that analysts may want to explore weighting based on considerations of survey quality—for example, underweighting countries with less-than-complete coverage or that use quota samples or substitution. Similarly, analysts may consider running models multiple times, iteratively dropping a single country from the analysis, to see if results are robust. Such approaches provide an appealing alternative to suggestions of simply throwing out data from countries whose samples may be less than ideal—countries that tend to be disproportionately non-Western.

**Mixed methods**

There is perhaps no greater potential for decolonizing methodology than through the appropriation of mixed methods in secondary cross-national work. Increasingly, survey programs, including the ESS and the Comparative Survey of Electoral System, collect country-specific contextual data to supplement the survey data. Van de Vijver and Chastiotis (2010) provide examples of effective integration of cross-national survey data with other forms of data collection, arguing that the potential impact of combining qualitative and quantitative insight into cross-national and cross-cultural studies is vast and largely unexplored (although analysts should note the unique challenges of cross-national qualitative research—e.g. Mangen, 1999). Ali (2011) provides an example here, as she carried out cognitive interviewing following the main cross-national data collection (rather than beforehand, as in the earlier example). Ali shows that respondents in Bangladesh simply did not interpret questions in the WVS on the meaning of democracy in the way that had been assumed by the central coordinating team.

**Conclusion**

As we noted in 2005, the increasing globalization of survey research has not necessarily meant the spread of an identical product across the globe—and for good reason. While survey methods and standards were, in many cases, developed in the Western world, their spread to non-Western countries has invited further consideration of their applicability. This piece has highlighted the role that survey bodies themselves have taken in establishing standards that more effectively integrate an increasingly diverse range of countries, as well as the potential for secondary-data analysts to contribute to the cross-national knowledge base as it becomes more heterogeneous.

In discussing both survey organizations and secondary analysts, this article has highlighted various ways in which comparative survey research methodology has been partly decolonized in recent years, and the ways in which it could be in the future. By this we mean that methodological practice, initiatives, and insights have been produced or influenced by research teams and research findings from around the world, not solely a small group of Western researchers exporting a narrow list of Western ideas.
To-date, decolonization has touched all aspects of the survey process. At the organizational level, non-Western research teams run a number of regional survey series, such as the Afrobarometer and Latinobarómetro. With regard to data collection, the ESS has effectively privileged both strict standardization in sampling procedures and necessary country variation in sample design, thus acknowledging that high-quality cross-national work requires local variation and insight. Local documentation of data collection processes also provides countries with some latitude in reporting those issues that may be unique but still relevant to the data collection process.

Privileging local understandings of more global concepts has also been shown to be crucial for ensuring measurement equivalence in cross-national surveys. Cognitive interviews provide one excellent way forward here, as such interviews provide the opportunity for those who are potentially ‘methodologically colonized’ to explicitly voice their own intellectual frameworks and measure them against Western methodologies.

Analysts provide much potential for further exploring the challenges of cross-national research, especially as it broadens into non-Western territory. For example, there is further scope for doing comparative research by merging data from different single-country sources, rather than relying on large-scale cross-national surveys. New models of meta-studies have also highlighted the need for analysts to use different control variables in regression models for different countries (e.g. Duch et al., 2010). Thus model specification does not need to be a uniform mold applied to all but can be driven by more country-specific considerations. Similarly, work on measurement modeling is beginning to explore how different countries may require different measurement models to ensure equivalence across countries. In all cases, input from local experts can help to ensure that differences between countries are ‘real’ and not methodological artifacts.

The continued globalization of public opinion research, in many ways, relies on the continued decolonization of survey methodology. Methodological advances such as those discussed above have undeniably strengthened the enterprise of comparative research by allowing a richer set of understandings to emerge from a broader set of contexts.

Notes
1. For example, see the section on ‘Analyzing Data’ in Harkness, Braun, Edwards, et al. (2010).
3. The two-step approach can also be considered a type of meta-analysis of regressions from a series of single-country studies, with the benefit of control over the initial estimates being analyzed at the second level.

Notes on contributors
Shawna N. Smith is a doctoral candidate in sociology at Indiana University—Bloomington. Her research examines cross-national differences in regulatory policy, from passage to implementation to efficacy, and cross-national methodology.

Stephen D. Fisher is a University Lecturer in Political Sociology at the University of Oxford. His research interests are in social attitudes, political behaviour and quantitative research methods.
Anthony F Heath holds professorial appointments at the Universities of Manchester and Oxford and is a Fellow of the British Academy. His research focuses on ethnicity and social inequality, and he has carried out a number of cross-national studies (including work on South Asia). He has published extensively on survey methodology and has also written a number of commissioned reports for government and international bodies including UNDP, OECD, the Cabinet Office, the Department for Work and Pensions and the Department for Communities and Local Government.

References


