Annu. Rev. Polit. Sci. 2005. 8:297–333
doi: 10.1146/annurev.polisci.8.090203.103000

First published online as a Review in Advance on Jan. 24, 2005

THE GLOBALIZATION OF PUBLIC OPINION RESEARCH

Anthony Heath, Stephen Fisher, and Shawna Smith
Department of Sociology, University of Oxford, Oxford OX1 3UQ, United Kingdom; email: anthony.heath@sociology.ox.ac.uk, stephen.fisher@sociology.ox.ac.uk, shawna.smith@sociology.ox.ac.uk

Key Words survey methodology, cross-national research, comparative research, survey history

■ **Abstract** As globalization has opened up channels of communication between different countries and increased interest in cross-national analysis, public opinion survey research has expanded its reach in the world. This article examines both the breadth and the depth of the globalization of public opinion research. First, we discuss the growth of cross-national surveys such as the World Values Survey, the International Social Survey Program, the European Social Survey, and the various Global Barometers. We then turn to the issues of data quality and comparability. Has the globalization of survey research meant the spread of a standard "product" of known and equivalent quality to diverse countries? Can survey research in diverse countries and contexts deliver meaningful comparisons of public opinion? Has globalization led to the dominance of an intellectual framework and set of assumptions that may not be quite appropriate outside their original homes? Finally, the article suggests a new standard for "grading" cross-national programs of survey research, inspired by debates in evidence-based medicine.

INTRODUCTION

If globalization is "the emergence of a global culture system. . .brought about by a variety of social and cultural developments" (Marshall 1998, p. 258), these social and cultural developments connote a number of processes specific to public opinion and survey research. For one, the greater ease and speed of communication between countries, seen in the close collaboration between national teams, makes these cross-national programs of survey research feasible. Another connotation of globalization is the export of Western technology and practices to less developed countries, seen in the export of survey research methodologies across the globe. And a third is the penetration of global brands to markets around the world. This article explores the extent to which a global "brand" or "product" of public opinion research has spread across the world, and what the current limitations and intellectual assumptions of such a brand might be.

We begin with a brief history of the spread of public opinion surveys, particularly noting growth among commercial, government, and academic surveys, and most specifically the growth of cross-national survey endeavors, around the world. We then turn to the issues of data quality and comparability. Has the globalization of survey research meant the spread of a standard "product" of known and equivalent quality to diverse countries? Can survey research in diverse countries and contexts deliver meaningful comparisons of public opinion? And finally, has globalization led to the dominance of a particular intellectual framework and set of assumptions that may not be quite appropriate outside their original homes?

In this review, we focus largely on academic survey research rather than on government or commercial opinion polls, and we focus particularly on the widely used and highly regarded cross-national survey programs. However, a great deal of what we say will also apply to academic surveys (and indeed to government and commercial ones) that are not part of these cross-national programs.

A BRIEF HISTORY

Commercial Polls

Although roots of the survey method can be traced back to the nineteenth century (see, e.g., Marsh 1982, Converse 1987), the start of systematic public opinion research is usually taken to be the work of George Gallup in the United States during the interwar years (Worcester 1987, Bulmer 1998, Rose & Osborne 1999). Although there were opinion polls before Gallup's, his contribution was the use of systematic samples of the population:

> If a sample is accurately selected, it represents a near replica of the entire population. It is a miniature electorate with the same proportion of farmers, doctors, lawyers, Catholics, Protestants, old people, young people, businessmen, laborers and so on, as is to be found in the entire population. (Gallup 1948, pp. 22–23)

The key features of survey research established by Gallup—and followed ever since—were the use of random samples from a defined population (typically quota rather than probability samples), the use of standardized "closed" questions to measure subjective attitudes and demographic characteristics of respondents, the administration of face-to-face surveys by trained interviewers, and the quantitative analysis of the results. The focus of these early surveys was on political attitudes and behavior; most famously, Gallup successfully predicted Roosevelt's win in the 1936 presidential election (in contrast to the *Literary Digest*, whose prediction of an Alf Landon victory relied on a nonrepresentative sample) (Rose & Osborne 1999). Gallup exported his approach and methods to Britain in 1937 and France in 1938, and commercial public opinion research rapidly spread during the war years to other wealthy industrialized countries, reaching Australia, Canada,

Denmark, Switzerland, the Netherlands, West Germany, Finland, Norway, and Italy by 1946 (Worcester 1987). In nearly all of these countries, Gallup polls (and others) continue to thrive, although telephone interviewing has gradually replaced face-to-face interviews, mainly owing to cost concerns.

A continuing theme has been the political opinion poll, for which commercial sponsors (such as the mass media) can readily be found. By the beginning of the 1980s, commercial political polls were being conducted in nearly all Western democracies and in a growing number of Communist and less-developed nations (Crespi 1989). Today that number is even larger; Gallup's most recent survey, The Voice of the People (http://www.voice-of-the-people.net), drew respondents from 51 countries in 2003, including a number of Middle Eastern, West Asian, and African countries. It is now rare to have a general election in a democracy without a pre-election poll [although, as Smith (2004) notes, bans on polls during the immediate run-up to an election still exist in some countries].

Government-Sponsored Polls

A second broad category of polls, government-sponsored surveys, began to grow at around the same time that Gallup's methods were being exported. Government polls, however, were designed to meet governments' needs for knowing about (and perhaps influencing) their citizens, and had the potential to be somewhat less benign than commercial polls. One of the earliest examples is the "survey" of each British county demanded by the Defense of the Realm Act of April 1798, conducted by the British government to ascertain the willingness of the (male) public to fight against the French in the event of an invasion (Colley 1992). More contemporarily, however, Samuel Stouffer and his team in the Research Branch of the Information and Education Division of the U.S. War Department conducted surveys of the U.S. armed forces between 1942 and 1945. These studies, collectively known as The American Soldier, used polling techniques to examine the attitudes and morale of soldiers, injecting for the first time social psychological questions and questioning subjective opinions and preferences (Stouffer et al. 1949a,b, 1950; Hovland 1950; Kuechler 1998; Rose & Osborne 1999). Interestingly, nondemocratic governments also have a history of carrying out opinion polls. When Khrushchev came to power in the Soviet Union, his criticisms of the lack of reliable information on his publics led to the formation of the Institute for Concrete Social Research and the appointment of Boris Grushin as the head of a center for studying public opinion in 1969 (Worcester 1987). And in the former Czechoslovakia, although the previously established Institute for Public Opinion Research was abolished and condemned as "bourgeois pseudoscience" by the Communist Party in 1948, it was re-established in 1965 within the newly formed Institute of Sociology. As Otava (1988) notes, however, the "empirical" results of these studies were not always used in good faith:

> [I]n the first issue of the *Sociological Journal* for 1987, a regular feature entitled "Empirical Surveys" presented the results of an investigation called

> "Public Opinion on Questions of War and Peace". . . . It begins with the sentence: "The empirical results confirm a positive evaluation of the foreign-policy orientation of the USSR. The opinion that the USSR is working for peace was unanimous. . . ." (pp. 252–53)

At best, government polls can give citizens a voice—although at worst they can be a source of political manipulation. The growing need to obtain the assent of the governed in an advanced society means that such government surveys have also become more global. Most recently, for example, the Coalition Provisional Authority in Iraq commissioned public opinion surveys to help formulate its policy.

An additional weakness of government-sponsored opinion surveys is that they are rarely available for secondary academic analysis. One important exception, however, is the Eurobarometer (http://europa.eu.int/comm/public_opinion/) sponsored by the European Commission. In 1962, the then EEC (European Economic Community) commissioned the "Attitudes to Europe" survey of Germany, France, and the Benelux counties. This was a prelude to the Eurobarometer, a biannual survey that started in 1970 and has included all European member states since 1973. Its coverage has expanded as the European Union has grown, and it has recently been supplemented by the Candidate Countries Eurobarometer (CC-EB; http://europa.eu.int/comm/public_opinion/), which was launched in 2001 as a replacement for the Central and Eastern Eurobarometer. As both of these surveys have EU funding, their primary purpose has been meeting the needs of the EU; however, the Eurobarometers have also had substantial academic input, notably collaboration with the Inter-University Consortium for Political and Social Research at the University of Michigan and the ZUMA (Zentrum für Umfragen, Methoden und Analysen) Center for Survey Research and Methodology at the University of Mannheim. These two institutes have also made the data available for secondary analysis, and the data have been widely used in academic research (see Inglehart 1990, 1997).

Academic Surveys

The main focus of this review is on a third category of surveys, namely academic surveys, funded by research councils or charities and directed by independent research institutes. Some of the earliest academic surveys of public opinion were again in the United States but were based on local samples. The Columbia studies of Erie County, OH in 1940 and Elmira County, NY in 1948, funded initially by the Rockefeller Foundation and guided by Berelson and Lazarsfeld, examined individuals in these two communities to study how voters made up their minds during presidential election campaigns. Both studies were panel studies, and attention was given to several individual variables as well as contextual variables, such as the mass media (largely in the Erie study) and interpersonal associations (in Elmira) (Lazarsfeld et al. 1948, Rossi 1964, Berelson et al. 1954).

Published several months before the Elmira study, but examining the 1952 election, was another study of the American electorate conducted by the Survey

Research Center (SRC) at the University of Michigan. The SRC's Michigan Studies, later the American National Election Studies (ANES; http://www.umich.edu/~nes/), differed from the Columbia studies by using relatively small but nationally representative probability samples and by focusing on individual motivations for party preferences. Campbell developed the sociopsychological intellectual framework embodied by the Michigan Studies in the classic works *The Voter Decides* and *The American Voter* (Campbell et al. 1954, 1960). The Michigan Studies shortly became the gold standard in electoral research and have conducted studies of every presidential and congressional election since 1948. The National Science Foundation (NSF) took over funding the studies in 1977, formally establishing the ANES.

After their establishment in the United States, the Michigan Studies were exported, not unlike Gallup's export, via links with scholars in other countries. The model was first exported to Britain, with the famous collaboration between Donald Stokes (part of the *American Voter* team) and the British psephologist David Butler. They began with a 1963 pre-election survey and instituted panel studies of electoral behavior and opinion. This led to the classic work on British politics and public opinion, *Political Change in Britain* (Butler & Stokes 1974), and helped to establish the British Election Studies (BES; http://www.essex.ac.uk/bes/index.html/), in which panel research continues to play a large role. The Michigan Studies were also exported in rather similar fashion to India with a 1967 survey (Eldersveld & Ahmed 1978), and they established what remains a notable tradition of Indian electoral research.

Many other wealthy societies now have regular election surveys funded by national science foundations and intellectually independent of national governments or other interested bodies; however, only a few have long histories like the ANES or BES. France, for example, has still to acquire a series of this kind, and although a series of German election surveys can be constructed going back to 1949, it is not a formal series in the American or British sense. However, over the past 20 years or so there has been a rapid spread of election studies around the world, and the Comparative Study of Electoral Systems (CSES, http://www.umich.edu/~cses), a cross-national collaboration that includes a standardized "add-on" module to individual country election surveys, now covers more than 50 states. The first module, running from 1996 to 2001, was completed in more than 30 diverse countries, including the Ukraine, Israel, Korea, Thailand, and Peru.

In addition to the ANES, a second widely copied model has been the American General Social Survey (GSS; http://www.norc.uchicago.edu/projects/gensoc.asp/), which began in 1972, funded by the Russell Sage Foundation and the NSF. The GSS has a broader remit than the election survey and includes standardized, closed questions on social attitudes and values as well as political issues; the survey was annual until 1994 and has been biennial since. Like the ANES, it uses a nationally representative probability sample with face-to-face interviewing. The British Social Attitudes survey, which started in 1983, is largely the British equivalent of the GSS and holds a similar remit. In Germany, the Allgemeinen Bevölkerungsumfrage der

Sozialwissenschaften (ALLBUS), begun in 1980, also serves as a biennial survey of social behaviors and attitudes.

Collaboration between the ALLBUS and the GSS occurred as early as 1982. In 1984 these three social surveys, in conjunction with representatives from the Research School of the Social Sciences at the Australian National University, agreed to a further and more developed program of cross-national collaboration, eventually in the form of the International Social Survey Program (ISSP; http://www.issp.org). The founding countries—the United States, Britain, Germany, and Australia—devised the format of add-on modules of standardized closed questions, designed by the cross-national team, with identical wording in each participating country. The ISSP continues to conduct annual surveys, investigating a new topic each year. It now surveys 38 countries, predominantly Western but including Bangladesh since 1996 and South Africa since 2001, as well as Japan, Chile, Venezuela, and several former communist countries. Table 1 gives details of the countries that have been included in the ISSP (and in other major cross-national programs) since inception.

Most extensive of the academic survey programs has been the World Values Survey (WVS; http://www.worldvaluessurvey.org), which can make some claims to being global. It began in 1981 as the European Values Survey (EVS, http://www.europeanvalues.nl/), which covered 10 West European societies. The funding for the initial survey was provided by a private foundation, and the focus of the research was on changing moral and social values, particularly Christian values and "alternative" meaning systems (see Halman 2001). The EVS has now completed three rounds of surveys, the latest of which commenced in 1999, and continues to focus on values and value systems among European countries. The WVS grew out of the EVS; after the successful completion of the 1981 EVS, the survey was replicated in 12 non-European countries. Following the success of this wave, subsequent waves were conducted in 1990–1993 (42 countries), 1995–1997 (54 countries), and 1999–2001 (60 countries). Surveys have now taken place in almost 80 societies that represent all major regions of the world, although, as WVS documentation notes, illiterate rural respondents have been underrepresented. All WVS surveys are carried out in face-to-face interviews, using a standardized sampling universe of adult citizens aged 18 and over. Fieldwork in each country is typically supported by funding from that country, and although the WVS has a relatively stable questionnaire, since the 1990 survey, participants from all six continents have been involved in design, fieldwork, analysis, and interpretation.

Akin to the WVS in some respects (although consciously modeled on the Eurobarometer) are the instruments now known as the Global Barometers (http://www.globalbarometer.org). The New Europe Barometer (http://www.cspp.strath.ac.uk; formerly New Democracies Barometer) series was founded by Richard Rose at Strathclyde University "to monitor mass response to the transformation of polity, economy and society in post-Communist countries." The study has regularly surveyed the eight new EU countries and three applicant countries (Croatia, Bulgaria, and Romania), and has conducted intermittent surveys in Serbia, Moldova, and, for

TABLE 1 Country coverage in cross-national research programs

	ISSP[a]	WVS/EVS[b]	Global Barometers[c]	CSES[d]	ESS[e]	Eurobarometers[f]
Albania		WVS: '95				CC-EB: '91–'96
Algeria		WVS: '00				
Argentina		WVS: '81, '90, '95, '00	Lat: '88, '95–'04			
Armenia		WVS: '95				CC-EB: '92–'96
Australia	'85–'88, '90–'02	WVS: '81, '95		'96		
Austria	'85–'89, '91–'95, '98–'02	WVS: '90 EVS: '90, '99	NE: '91, '98, '04		'02	EB: '94–'04
Azerbaijan		WVS: '95				
Bangladesh	'96–'97, '99–'00	WVS: '95, '00				
Belarus		WVS: '90, '95 EVS: '99	NE: '92–'93, '95, '98, '04	'01a		CC-EB: '92–'96
Belgium	'00, '02	WVS: '81, '90 EVS: '81, '90, '99		'99	'02	EB: '74–'04
Bolivia			Lat: '96–'04			
Bosnia-Herzegovina		WVS: '95				
Botswana			Afro: '99, '03, '05			
Brazil	'00–'02	WVS: '90, '95	Lat: '88, '95–'04			
Bulgaria	'92–'00, '02	WVS: '81, '90 EVS: '99	NE: '91–'95, '98, '01, '04	'01b		CC-EB: '90–'97, '01–'03

(*Continued*)

TABLE 1 (*Continued*)

	ISSP[a]	WVS/EVS[b]	Global Barometers[c]	CSES[d]	ESS[e]	Eurobarometers[f]
Canada	'92–'01	WVS: '81, '90, '00		'97		
Cape Verde			Afro: '02, '05			
Chile	'98–'02	WVS: '90, '95, '00	Lat: '88, '95–'04	'99		
China		WVS: '90, '95, '00	EAsia: '94, '02			
Colombia		WVS: '95	Lat: '96–'04			
Costa Rica			Lat: '96–'04			
Croatia		WVS: '95 EVS: '99	NE: '92–'93, '95, '98, '04			CC-EB: '95–'96
Cyprus	'96–'02					CC-EB: '01–'03
Czech Republic	'93–'02	WVS: '90, '95 EVS: '99	NE: '91–'95, '98, '01, '04	'96	'02	CC-EB: '90–'97, '01–'03 EB: '02
Denmark	'98–'02	WVS: '81, '90 EVS: '81, '90, '99		'98	'02	EB: '74–'04
Dominican Republic		WVS: '95	Lat: '02, '04			
Democratic Republic of the Congo		WVS: '95				
Ecuador			Lat: '96–'04			
Egypt		WVS: '00				
El Salvador		WVS: '95	Lat: '96–'04			
Estonia		WVS: '90, '95 EVS: '99	NE: '98, '01			CC-EB: '91–'97, '01–'03

Finland	'00–'02	WVS: '81, '90, '95 EVS: '81, '90, '99			'02	EB: '93–'04
France	'96–'02	WVS: '81, '90 EVS: '81, '90, '99		'02	'02	EB: '74–'04
Georgia		WVS: '95				CC-EB: '92, '94–'96
Germany[g]	'85–'02	WVS: '81, '90, '95 EVS: '81, '90, '99	NE: '04	'98, '02	'02	EB: '74–'04
Great Britain	'85–'02	WVS: '81, '90, '95 EVS: '81, '90, '99		'97	'02	EB: '74–'04
Ghana		WVS: '95	Afro: '97, '99, '02, '05			
Greece		EVS: '99			'02	EB: '80–'04
Guatemala			Lat: '96–'04			
Honduras			Lat: '96–'04			
Hong Kong			EAsia: '94, '01	'98, '00		
Hungary	'86–'02	WVS: '81, '90, '95 EVS: '81, '99	NE: '91–'95, '98, '01, '04	'98, '02	'02	CC-EB: '90–'97, '01–'03 EB: '02
Iceland		WVS: '81, '90 EVS: '81, '90, '99		'99		EB: '03
India		WVS: '90, '95, '00				
Indonesia		WVS: '00	EAsia: TBC			
Iran		WVS: '00				

(*Continued*)

TABLE 1 (*Continued*)

	ISSP[a]	WVS/EVS[b]	Global Barometers[c]	CSES[d]	ESS[e]	Eurobarometers[f]
Ireland	'86–'96, '98–'00, '02	WVS: '81, '90 EVS: '81, '90, '99		'02	'02	EB: '74–'04
Israel	'89–'91, '93–'94, '96–'02	WVS: '00		'96	'02	
Italy	'85–'01	WVS: '81, '90 EVS: '81, '90, '99			'02	EB: '74–'04
Japan	'93–'02	WVS: '81, '90, '95, '00	EAsia: TBC	'96		
Jordan		WVS: '00				
Kazakhstan						CC-EB: '94–'96
Kenya			Afro: '03, '05			
Latvia	'95–'02	WVS: '90, '95 EVS: '99	NE: '98, '01			CC-EB: '91–'97, '01–'03
Lesotho			Afro: '00, '03, '05			
Lithuania	'94	WVS: '90, '95 EVS: '99	NE: '98, '01	'97		CC-EB: '91–'97, '01–'03
Luxembourg		EVS: '99			'02	EB: '74–'04
Macedonia		WVS: '95				CC-EB: '93–'96
Malawi			Afro: '99, '03, '05			
Mali			Afro: '01, '02, '05			
Malta		EVS: '99				CC-EB: '01–'03
Mexico	'00, '02	WVS: '81, '90, '95, '00	Lat: '95–'04	'97, '00		
Moldova		WVS: '95	NE: '04			CC-EB: '92
Mongolia			EAsia: TBC			

Montenegro		WVS: ’95, ’00	NE: ’98			CC-EB: ’96
Morocco		WVS: ’00				
Mozambique			Afro: ’02, ’05			
Namibia			Afro: ’99, ’02, ’03, ’05			
Netherlands	’87–’89, ’91, ’93–’95, ’97–’00, ’02	WVS: ’81, ’90 EVS: ’81, ’90, ’99		’98	’02	EB: ’74–’04
New Zealand	’91–’02	WVS: ’95		’96, ’02		
Nicaragua			Lat: ’96–’04			
Nigeria		WVS: ’90, ’95, ’00	Afro: ’00, ’01, ’03, ’05			
Northern Ireland	’01–’02	WVS: ’81, ’90 EVS: ’81, ’90, ’99			’02	EB: ’75–’04
Norway	’89–’02	WVS: ’81, ’90, ’95 EVS: ’81, ’90		’97	’02	EB: ’90, ’92–’04
Pakistan		WVS: ’95, ’00				
Panama			Lat: ’96–’04			
Paraguay			Lat: ’95–’04			
Peru		WVS: ’95, ’00	Lat: ’95–’04	’00, ’01a		
Philippines	’91–’02	WVS: ’95, ’00	EAsia: ’02			
Poland	’87, ’91–’02	WVS: ’90, ’95 EVS: ’99	NE: ’91–’95, ’98, ’01, ’04	’97, ’01b	’02	CC-EB: ’90–’97, ’01–’03
Portugal	’97–’00, ’02	WVS: ’90 EVS: ’90, ’99		’02	’02	EB: ’85–’04

(*Continued*)

TABLE 1 (*Continued*)

	ISSP[a]	WVS/EVS[b]	Global Barometers[c]	CSES[d]	ESS[e]	Eurobarometers[f]
Puerto Rico		WVS: '95, '00				
Romania		WVS: '90, '95 EVS: '90, '99	NE: '91–'95, '98, '01, '04	'96		CC-EB: '91–'97, '01–'03
Russia	'91–'02	WVS: '90, '95 EVS: '99	NE: '04	'99		CC-EB: '90–'96
Senegal			Afro: '02, '05			
Serbia		WVS: '95, '00	NE: '98, '04			CC-EB: '96
Slovakia	'96, '98–'00, '02	WVS: '90, '95 EVS: '99	NE: '91–'95, '98, '01, '04			CC-EB: '92–'97, '01–'03
Slovenia	'91–'02	WVS: '90, '95 EVS: '90, '99	NE: '91–'95, '98, '01, '04	'96	'02	CC-EB: '92–'97, '01–'03
South Africa	'86, '01	WVS: '81, '90, '95, '00	Afro: '94, '95, '97, '98, '00, '02, '05			
South Korea		WVS: '81, '90, '95, '00	EAsia: '94, '96, '98, '99, '01	'00		
Spain	'92–'02	WVS: '81, '90, '95, '00 EVS: '81, '90, '99	Lat: '01–'04	'96, '00	'02	EB: '85–'04
Sweden	'92, '94–'00, '02	WVS: '81, '90, '95, '00 EVS: '81, '90, '99		'98	'02	EB: '94–'04
Switzerland	'87, '93, '98–'02	WVS: '90, '95 EVS: '90		'99	'02	

Taiwan	’02	WVS: ’95	EAsia: ’94, ’98, ’01	’96	
Tanzania		WVS: ’00	Afro: ’01, ’03, ’05		
Thailand			EAsia: ’01	’01a	
Turkey		WVS: ’90, ’95, ’00 EVS: ’99			CC-EB: ’01–’03
Uganda			Afro: ’00, ’02, ’05		
Ukraine		WVS: ’95 EVS: ’99	NE: ’92–’93, ’95, ’98, ’04	’98	CC-EB: ’92–’96
United States	’85–’02	WVS: ’81, ’90, ’95, ’00		’96	
Uruguay		WVS: ’95	Lat: ’88, ’95–’04		
Valencia		WVS: ’95			
Venezuela	’99–’00	WVS: ’95, ’00	Lat: ’95–’04		
Vietnam		WVS: ’00			
Zambia			Afro: ’93, ’96, ’99, ’03, ’05		
Zimbabwe			Afro: ’99, ’04, ’05		

[a]Annual waves, reported 1985–2002; date listed is date of wave, not necessarily of fieldwork [source: http://www.issp.org (1985–2000) and http://www.gesis.org/en/data_service/issp/index.htm (2001–2002)].

[b]WVS, four waves: 1981–1984, 1990–1993, 1995, 2000; EVS, three waves: 1981–1984, 1990–1993, 1999 (source: http://www.worldvaluessurvey.org).

[c]Afrobarometer (Afro): round 1, 1999–2001; round 2, 2002–2004; round 3, 2005 (planned); plus additional time series in some countries (source: http://www.afrobarometer.org). East Asia (EAsia): no systematic waves (source: http://www.eastasiabarometer.org). Latinobarómetro (Lat): 1988 pilot, annual since 1995 (except 1999) (source: http://www.latinobarometro.org). New Europe (NE): annual 1991–1995, 1998 (as New Democracies Barometer), 2001, 2004 (source: http://www.cspp.strath.ac.uk).

[d]Module 1: 1996–2001a; Module 2: 2001b–2005 (source: http://www.umich.edu/~cses/).

[e]Round 1, 2002 (source: http://www.europeansocialsurvey.org).

[f]Candidate Countries Eurobarometer, formerly Central and Eastern Eurobarometer (CC-EB): annual 1990–1997, biennial 2001–2003. Eurobarometer (EB): biennial 1985–2004 (sources: http://europa.eu.int/comm/public_opinion/ and http://www.gesis.org/en/data_service/eurobarometer/).

[g]Prior to 1990, typically West Germany only.

comparison, Austria and reunified Germany. The Latinobarómetro (http://www.latinobarometro.org) was established next in 1995, following a four-country pilot study in 1988. It initially covered eight countries in Latin America (excluding the Caribbean) but expanded to 17 countries in 1996. The Latinobarómetro is designed as a time series with a rotation of topics included in the survey, such as attitudes toward international trade and the environment, patterns of political participation, and gender and discrimination. Funding for this series has been provided by the Corporación Latinobarómetro, a private, nonprofit initiative, and a large international board of notable academics oversees the project. The Afrobarometer (http://www.afrobarometer.org) completed its first round of studies covering 12 countries in 2001 and commenced a second wave in 2002. The Afrobarometer has been funded by a number of African and non-African governmental agencies, including the NSF and the U.S. Agency for International Development. From round two, an identical survey instrument has been used in all countries, and sample sizes within countries range from 1200 to 2400. Finally, the East Asia Barometer (http://www.eastasiabarometer.org), the newest of the Global Barometers, began surveys in late 2001 with funding from the Ministry of Education of the Republic of China. To date, surveys have been conducted in Taiwan, South Korea, Hong Kong, Thailand, the Philippines, mainland China, and Mongolia, with additional fieldwork in Indonesia and Japan pending. However, no results are yet available publicly, nor is technical information about the fieldwork. A new five-nation "State of Democracy in South Asia" survey (http://www.lokniti.org/projects.htm#sdsa), organized by the Center for the Study of Developing Societies in Delhi, is now under way; this survey could be thought of as a South Asian Barometer, and may perhaps be the next to join the conglomerate of Global Barometers.

Whereas most cross-national surveys, such as the ISSP and WVS, are administered by different organizations in different countries with different sampling and fieldwork methodologies, the European Social Survey (ESS; http://www.europeansocialsurvey.org/) was established in 2002 with an emphasis on methodological rigor and uniformity. Twenty-two nations participated in the first round, namely the then 15 member states of the EU plus four accession states (the Czech Republic, Hungary, Poland, and Slovenia) and three non-EU members (Israel, Norway, and Switzerland). All ESS surveys are based on face-to-face probability samples, and detailed rules are provided in all participating countries, including study-wide targeted response rates of 70% and strict rules regarding sampling (Lynn et al. 2004). Overall the ESS aims to be a cross-national social survey that achieves "uniform methodological standards that make it at least as rigorous as the very best national surveys within Europe."[1]

The most recent foray into cross-national survey research has been the Pew Global Attitudes Project (http://people-press.org/pgap), started in 2002. This

[1]R. O'Shea, C. Bryson, R. Jowell. Undated. *Comparative attitudinal research in Europe*. http://www.europeansocialsurvey.org/

survey, funded by the Pew Charitable Trusts, lacks the academic oversight of previously discussed surveys, but is far reaching in coverage. The 2002 survey included 44 nations and involved more than 38,000 people. Of particular interest in this first wave of studies were global attitudes about America, particularly in light of the 9/11 attacks, and particularly among Muslim communities. The Global Attitudes Surveys mark some of the first large-scale survey experience in the Middle East, including samples from Egypt, Pakistan, Jordan, Lebanon, Turkey, and Uzbekistan. Subsequent waves have included individuals from 50 populations.

Table 2 summarizes the characteristics of these main cross-national survey programs.

Of course there have also been many academic social surveys that have been part of neither a regular national series nor a cross-national survey program. Particularly notable is Almond & Verba's (1963, 1989) pioneering five-nation civic culture study of 1959, which covered the United States, Britain, Italy, Germany, and Mexico. The civic culture studies were unusual for their extensive use of open-ended questions, and this has not become the norm in cross-national research. The eight-nation *Political Action* surveys, conducted between 1973 and 1975 in Britain, West Germany, the Netherlands, Austria, the United States, Italy, Switzerland, and Finland, were similarly purpose-built; they examined forms of political participation in industrialized countries (Barnes & Kaase 1979). It is, however, beyond the scope of this paper to document or discuss these.

As this section has shown, the expansion of survey research throughout the world has been to date rapid and relatively far reaching. Whereas in the 1950s and 1960s academic public opinion research was largely relegated to a handful of wealthy, industrialized Western countries, by the early 1980s this tradition had grown to include most of North America and Europe, although the majority of these countries were still predominantly wealthy. Coverage of Latin America and Asia began during the 1990s, and systematic public opinion research has been undertaken in Africa and the Middle East only over the past five years. Thus, although public opinion research has clearly been affected by the pursuits of globalization, only very recently could it be considered a global phenomenon.

Even now it is not yet completely global. Certain areas of the world, particularly the Middle East and sub-Saharan Africa, are still underrepresented within large cross-national programs, and illiterate and rural populations are underrepresented in nearly all parts of the world. And although recognition of the value of public opinion research and cross-national comparisons has grown since the 1950s, Smith (2004) notes that independent surveys are still banned in a number of nations (including Burma, Cuba, Laos, North Korea, and Turkmenistan), and in other countries (such as China, Venezuela, and Iran), polling topics and/or the publication of results are restricted (Rohme 1997, Spangenberg 2003).

This section has focused on the breadth of cross-national research, but understanding the extent to which public opinion research has been globalized also requires an understanding of the depth of the research. In other words, how does the quality of public opinion research in newly defined markets compare to standards

TABLE 2 Features of cross-national survey programs[a]

Series	Inception[b]	Countries covered[c]	Frequency	Survey type	Mode of data collection	Available data	Website
Comparative Study of Electoral Systems (CSES)	1996	36	Module every 5 years	Module	Face-to-face, telephone, and self-completion	Public archives	www.umich.edu/~cses
Eurobarometers:							
Candidate Countries Eurobarometer (CC-EB)	1990	24	Annual	Standalone	Face-to-face	Public archives	http://europa.eu.int/comm/public_opinion/
Eurobarometer	1973	19	Bi-annual	Standalone	Face-to-face[d]	Public archives	http://europa.eu.int/comm/public_opinion/
European Social Survey (ESS)	2002	22	Biennial	Standalone	Face-to-face	Public archives	www.europeansocialsurvey.org
European Values/World Values Surveys (EVS/WVS)	1981/1983	33/77	~5 years	Standalone	Face-to-face	Public archives	www.worldvaluessurvey.org www.europeanvalues.nl/index2.htm
International Social Survey Program (ISSP)	1985	39	Annual	Module	Face-to-face and self-completion	Public archives	www.issp.org

Gallup Int'l Voice of the People Survey	2002	60	Annual	Standalone	Telephone and face-to-face	Tables only	www.voice-of-the-people.net/
Global barometers:							
Afrobarometer	1999	16	Annual	Standalone	Face-to-face	Public archives	www.afrobarometer.org
East Asia Barometer	2001	9	Annual	Standalone	Face-to-face	None to date	www.eastasiabarometer.org
Latinobarómetro	1995	19	Annual	Standalone	Face-to-face	Tables only	www.latinobarometro.org
New Europe Barometer	1991	18	Annual	Standalone	Face-to-face	Tables only	www.cspp.strath.ac.uk/
Pew Global Attitudes Survey	2002	49	Annual	Standalone	Telephone	Tables only	http://people-press.org/pgap

[a]This table is an expansion of the table presented by Norris (2004).

[b]In all cases but the CC-EB and the New Europe Barometer, pilot studies and forerunners (such as the European Community Study which preceded the Eurobarometer) have been excluded. As the CC-EB replaced the Central and Eastern Eurobarometer (CEEB), the CEEB's date of inception is listed; similarly the New Democracies Barometer preceded the New Europe Barometer and thus its date of inception is listed.

[c]Number of countries included in at least one survey.

[d]Interviews in Sweden for survey 42 were carried out by telephone.

for research in established markets; or, more importantly, how comparable are the surveys undertaken by cross-national survey programs? The following section explores these methodological issues.

METHODOLOGICAL ISSUES IN CROSS-NATIONAL RESEARCH

Although there is no strict division, cross-national survey research can be thought of as being affected both by general methodological issues relevant to any survey (survey quality) and by the comparability of responses from different countries (equivalence of meaning). The spread of public opinion research to a greater number and diversity of countries almost certainly entails greater problems of equivalence of meaning and consistency of survey quality. Harkness (1999) argues that, in the cross-national context, discussions of quality are rare compared with discussions of equivalence of meaning. This section examines both issues. We begin with survey quality because it is, in a sense, prior.

We recognize that our terminology here is somewhat different from the usual one, which would treat issues such as equivalence of meaning as one aspect of data quality. We draw a pragmatic distinction between the quality of the surveys implemented by the separate national teams and the problems of the questionnaire content designed by the central coordinating committee of the cross-national program.

Quality Issues

Data quality is an issue with all surveys, including those in the affluent democratic societies, but there are particular issues with the extension of survey research to countries outside the traditional core. The GSS may be considered the gold standard of survey research, but the extension of such research has by no means spread the application of the same survey methods used in the GSS. In short, surveys have not been a standard product exported across the world.

There are many reasons why different survey methods are used in different countries, even if they are part of the same cross-national project. Different countries and survey research institutions have different kinds of sampling frames, legislation regarding survey practice, traditions of how to pick samples, technical expertise, methods of recruiting and training interviewers, access to experienced interviewers and supervisors, access to computer equipment for computer assisted personal/telephone interviewing (CAPI/CATI), and practices of using call-backs and/or conversion techniques for nonrespondents and refusals. Populations differ in the extent of their experience with survey research and their levels of willingness to participate in it.

Groves (1987, 1989) distinguishes the following components of survey quality: coverage error, nonresponse error, sampling error, and measurement error. We examine each of these in turn.

COVERAGE Coverage error refers to "the discrepancy between sample survey results and the results of a full enumeration of the population under study which arises because some members of the population are not covered by the sampling frame" (Groves 1987, p. S159). Lynn et al. (2004) also argue that equivalent study populations are one of two fundamental criteria for comparable sample design (the other being similar precision of sample-based estimates, which we examine below). Usually in public opinion research the population is taken to be the adult population, although the definition of when adulthood starts is far from clear cross-nationally. The ESS, for example, takes the target population to be all residents in private households 15 years or older, whereas the sampling universe for the WVS is all citizens aged 18 and older.

The true level of coverage, however, tends to depend on the nature of the sampling frame and on the cost of reaching some of the groups included in the frame. Sample frames might be drawn from complete records of the resident population (e.g., the Danish Central Person Register), lists of households (e.g., the SIPO database in the Czech Republic), registers of addresses (e.g., the Postcode Address File in the United Kingdom), electoral registers (e.g., in India), or no records at all (e.g., in Bosnia-Herzegovina) (Lynn 2003a,b). Complete registers of the population are valuable as sample frames because they enable researchers to pick equal-probability samples; but even when they exist, the availability of registers to academic researchers varies, and they are sometimes restricted for use by the national statistics institute only. Coverage problems can also arise from parochial practices. Park & Jowell (1997) note that in the 1995 ISSP, five countries (mostly European) imposed an upper age cut-off for the sample at 74. Institutional populations in many countries (e.g., the United Kingdom) are also often overlooked in sampling designs.

Problems of coverage have been growing even in the affluent West, where cost issues have driven the growth of telephone interviewing and the consequent restriction of the sample to individuals with access to telephones. Cost has also tended to limit coverage of face-to-face surveys to accessible areas, e.g., exclusion of the highlands north of the Caledonian Canal in many British surveys because of the expense of reaching potential respondents.

Not surprisingly, coverage is likely to be a particular issue in less developed countries with less urbanized populations and with greater difficulties of access to much of the rural population. Inglehart (1997), writing about the 1991 WVS, is admirably frank about the problems:

> In Chile, the sample covers the central portion of the country, which contains 63% of the total population; the income level of this region is about 40% higher than the national average. In Argentina, sampling was limited to the urbanized central portion of the country. . .which also has above-average incomes. In India, the sample was stratified to allocate 90% of the interviews to urban areas and 10% to rural areas, and to have 90% of the [interviews] with literate respondents (who are slightly less than 50% of the population). In Nigeria,

> the fieldwork was limited to urban areas plus a sample of rural areas within 100 kilometers of an urban center. In China the sample is 90% urban. The samples have been weighted accordingly to make the samples replicate the national population parameters more closely. (p. 346)

The Afrobarometer Sampling Protocol (2002, http://www.afrobarometer.org/SamplingProtocol.pdf) also notes that areas experiencing armed conflict are excluded from sampling, and case-by-case judgments are made as to whether to include areas experiencing political unrest. The Pew Global Attitudes survey restricted areas even further, limiting its surveys in some countries to major cities—e.g., Luanda in Angola and Cairo in Egypt.

The extent to which weighting can deal with the problem of lack of coverage is an unresolved (and in many cases perhaps an irresolvable) question. The crucial issue is whether the opinions of the nonsampled population are in fact similar to those of their notional equivalents in the sampled population. For example, are rural Nigerians who live more than 100 km from an urban center similar in their opinions to rural Nigerians who live within 100 km? In other words, does public opinion show an interaction effect between rural residence and distance from an urban center?

This is in principle a researchable question. For example, in India, the National Election Survey is a probability sample of the adult population (drawn from the electoral register and hence with some problems of coverage); it could be used to explore whether such interactions are present and whether weighting of India's WVS survey, with its more limited coverage, would be sensible. However, in general, such research has not been carried out, and users of cross-national datasets need to be warned of the potential problems of differing coverage.

NONRESPONSE Nonresponse error has been a major concern of Western survey research organizations, and there appears to be a widespread decline in response rates in affluent, urbanized countries. For example, according to the GSS website (http://www.norc.uchicago.edu/projects/gensoc3.asp), GSS response rates have fallen from just under 80% in the 1980s to 70% in 2002. Even the most methodologically rigorous of the cross-national programs, the ESS, with a target response rate of 70% for all countries, experienced considerable variation in response rates during its first round; five countries had response rates below 50%, the lowest being 33.5% for Switzerland (Lynn et al. 2004). Add-on modules favored by the ISSP may well have lower response rates than the main survey.

The identification of low response rates is difficult because the globalization of survey research has not led to the globalization of good survey documentation. As Harkness (1999) points out, documentation about details such as response rates is often unavailable, or inaccessible without considerable detective work. We believe that it should be the responsibility of the cross-national programs themselves to provide this documentation for the participant surveys.

The problem is compounded further by differing usages of sampling methods. Cross-national programs often use quota samples, random-route methods, or

sampling methods that permit substitution—particularly in less developed countries, which have little infrastructure for survey research. In such cases, response rates cannot be calculated. Although the ESS strictly forbids both substitution and quota sampling, other survey programs continue to make use of these methods. During the 2001 ISSP, three countries reported using quota sampling at different stages, and 12 reported some level of substitution (Klein & Harkness 2003). However, as Groves (1987) points out, nonresponse bias is not simply a matter of nonresponse rates but also of the difference of means for respondents and nonrespondents; thus, bias is ultimately the key issue—not the response rate per se.

Checks on the size of the bias can sometimes be carried out directly for public opinion surveys—especially electoral surveys, for which the official records of aggregate vote will be available. Where an up-to-date and accurate census is available, the demographic profile of respondents can be checked against census figures. It is not known how nonresponse bias varies across countries (or indeed across surveys within a country), but clearly it could be a significant problem in cross-national research. Given the problem of declining response rates in the case of probability samples, methodologists should perhaps pay more attention to estimating the extent of nonresponse bias that arises from different sampling procedures.

SAMPLING In contrast to coverage error and nonresponse error, the principles of sampling error are well understood, at least in the case of probability samples. However, some issues in the spread of survey research need more investigation. The standard of academic acceptability has been a probability sample with a sample size of $\simeq$1000 (assuming population estimates with a confidence interval of $\pm$3% are required). Most cross-national surveys programs, including the EVS/WVS, the ISSP, the Eurobarometer, and the CC-EB, tend to target samples of this size.

There are several problems, however, with this standard. Clustering is often used (for cost reasons), and this tends to reduce the effective sample size. Many of the highest-quality surveys report design effects—the ratio of the variance of a variable to that which would have been obtained under simple random sampling (see Kish 1965 and Lynn et al. 2004 for more on design effects). Because a clustered design increases the variance, design effects are nearly always greater than one but tend to be relatively low in Western Europe. Modern software allows one to take the clustering into account in the analysis stage; however, because of ethical concerns about disclosure, it is becoming increasingly rare for the datasets released to investigators to identify the cluster of a particular respondent. Most analysts therefore tend to ignore the sampling error due to clustering, either because it is small or because they do not have the information to do so.

However, it is not clear that design effects are always small. They may generally be small in a relatively homogeneous West European society (e.g., in the ESS, the average predicted design effects were calculated as $\sim$1.5 in Great Britain, Germany, and France), but even in these countries we find that ethnic groups tended

to be geographically concentrated, and much larger design effects are associated with variables strongly related to ethnicity. This may also be true for ethnically diverse societies, such as India or many African countries, and may be both a methodological and a substantive problem. The precision of sample estimates may therefore vary across countries in unknown ways, and the substantive influence of local contexts on public opinion may be neglected. A practice that in the past was broadly acceptable in Western Europe may not be a good model for more heterogeneous societies.

It is also unclear that the standard sample of 1000 found in many cross-national programs is equally appropriate everywhere. This depends in part on the purpose of the survey. We are not always interested in simple population estimates of public opinion, as in the typical opinion poll of government popularity. We may also be interested in relations between variables; for example, in Northern Ireland, polarized between Catholics and Protestants, a larger sample size will be needed in order to provide sufficient power to determine the extent of difference between Catholic and Protestant public opinion. Divided societies need much larger sample sizes than homogeneous ones, and a target sample of 1000 may be quite misleading. The surveys may simply lack the power to detect differences in public opinion, and these differences may be the most important aspect of public opinion(s) in that country. In principle, there is no technical difficulty in designing appropriate sample sizes for diverse or divided societies. The Afrobarometer, for example, has set larger sample sizes ($N = 2400$) for those countries it considers "extremely heterogeneous," namely South Africa and Nigeria, and this must be considered the correct strategy.

MEASUREMENT Measurement error—the discrepancy between respondents' attributes and their survey responses—is perhaps the largest topic of all. It has been extensively studied, particularly in the United States by scholars such as Groves (see also Schumann & Presser 1981, Lyberg et al. 1997). Groves (1987) distinguishes the measurement errors that arise from "the influence of the interviewer, the 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" (p. S162). We deal with some aspects of the weakness of survey questions in the next section on equivalence of meaning, but a few observations relevant to the globalization of survey research may be in order.

It is perhaps most useful to distinguish between what we might call noise and bias. In countries with shorter traditions of survey research and a less well-trained field force, there may be greater interviewer variability in the conduct of the interview (e.g., in how strictly they adhere to the interviewer instructions), and there may be less supervision of interviewers to weed out "bad practice." (As with most issues that we cover, we must emphasize that there will also be important within-country variations—a cheaper survey in an affluent Western society may well involve a trade-off with measurement error.) This variability could lead to greater noise, which in turn could lead to weaker observed relationships between

variables. If we are simply interested in population estimates, this may not be important, since it may not affect the mean. However, noise clearly could be important if our interest is in studying relationships between variables, as in most academic research on public opinion. If we discover, for example, that social class is less strongly associated with socialist versus free-market values in developing countries than in developed ones, we cannot be sure whether this is because the values (or, indeed, social class) have been measured with more error or whether the relationship is actually weaker. Interviewer assessment of social class or self-coded measures of class, for example, will almost certainly have greater noise than more expensive office-coded measures based on details of occupation. This strikes at the heart of a great deal of cross-national research.

Bias is a separate matter. Whereas noise is of particular concern when one is looking at relationships between variables, bias becomes a central concern in estimating population means, proportions, or other quantities of interest. Differences in methods of survey administration, such as different modes or different sorts of interviewers, can lead to bias, but perhaps the main sources of bias arise from questionnaire content, to which we turn in the next section.

SUMMARY Globalization has not meant the spread of a standard social survey methodology across the world. The gold standard may be achieved by the GSS and some (but by no means all) European surveys with complete coverage, high response rates and low nonresponse bias, known and appropriate precision of sample estimates, and minimal measurement error; however, it is clear that survey research often falls short of this standard both within and between countries. Although we have highlighted the problems of coverage in developing countries, the wide variety of response rates—even in such a methodologically rigorous program as the ESS—indicates that major problems exist even in the developed world despite its relatively long tradition of survey research.

To be sure, the imposition of the same standard of survey methods in different countries, with their differing traditions of survey research, may not be possible or desirable. For example, different modes of survey administration (CATI, CAPI, etc.) may be appropriate in different countries with different histories and problems. Random-route or quota samples may be quite acceptable alternatives in a country where probability samples could not be implemented effectively, provided they do not involve excessive nonresponse bias.

However, the requisite information to enable one to judge the quality of surveys or the appropriateness of the method to the particular country is not always available cross-nationally. The documentation for cross-national survey research needs to be especially thorough but is rarely available (Harkness 1999). Requirements include (*a*) standard documentation on methods in each country; (*b*) meta-documentation for the program as a whole, highlighting and explaining the differences in the methods and contexts for the surveys; (*c*) translation notes, with guidance on the use of functional equivalence; and (*d*) indications of the implications for analysis. Perhaps the biggest problem is the difficulty of finding the technical details of the

surveys in order to make an informed judgement. The ESS is a model; its website offers information about most aspects of fieldwork, including response rates, as well as some assessment of measurement error. The ISSP, although it falls far short of the strict ESS methodology (likely because of a lack of central organization and funding), is to be commended for including regular reports monitoring the study (see Park & Jowell 1997, Sinnott 1998, Klein & Harkness 2003).

In summary, the consumer of cross-national research (whether the secondary data analyst or the reader) is often ignorant of the extent to which the quality of the "product" varies and how this might affect the results and interpretation.

Equivalence of Meaning

While survey design and implementation remain far from standardization, questionnaire content is much closer. Although some early surveys (such as Almond & Verba's civic culture project) made great use of open-ended questions, which allow respondents to articulate their opinions using their own concepts and language, usual survey practice has favored closed questions that can be asked in identical format in different surveys. However, these questions are often "decontextualized" in order to achieve identical wording.

Several different, albeit related, issues are relevant to the study of equivalence of meaning across countries. Perhaps most fundamentally, there may simply not be common concepts to measure (e.g., the concept of God may be specific to certain religious traditions). Second, there may be common concepts whose interpretation varies in different contexts. Third, poor translation may introduce errors. We deal with these issues in reverse order.

Translation of questions into different languages is inevitably fraught with problems, and in some surveys translation has not received enough attention. Sinnott (1998), for example, demonstrated that the Eurobarometer question on party identification had a systematically different meaning if it was derived from the English-language version rather than the French-language version of the original questionnaire. Whereas the English version asked the equivalent of "Are you close to a political party?" the French version asked, "Are you closer to one party than the others?" As Sinnott (1998) notes, "An individual who is only moderately or weakly aligned with a party could well answer no to the English version. . .and yes to the French version" (p. 631). And indeed, Sinnott's results reflect this disparity: More individuals answered yes to the relative (French) version of the question, and more answered no to the absolute (English) version.

The standard method of improving the quality of translation is to use "back translation"—the translation of a question from one language into another and then back again by a separate translator. By comparing the original with the doubly translated question, it should be clear whether there are any problems. However, Harkness & Schoua-Glusberg (1998) argue that in practical and theoretical terms (results, effort, costs, reliability, viability), back translation is one of the less recommendable procedures. Harkness (1999) describes it as merely a procedure for

checking translations. Warwick & Osherson (1973) similarly argue that back translation is a method of achieving linguistic equivalence and does not take account of more contextual factors.

Translation problems can often be solved, as for example in the Eurobarometer case described by Sinnott (1998), where a revision of the questionnaire appears to have led to greater comparability. They are probably not, in themselves, major sources of lack of comparability. Different interpretations of the same appropriately translated question can raise more difficult problems. For example, in its module on migration, the ESS has a question asking whether more individuals from "poorer countries outside Europe" should be allowed to enter. This is a question that can be asked in all countries with the same wording, but will it be interpreted in the same way in the different countries? In Israel, for example, migrants from poorer countries may be taken to mean Jewish migrants from North Africa and the Middle East, but in Britain the same phrase may be interpreted to refer to migrants from South Asia or the Caribbean. The question has been decontextualized in a way that permits standardization of wording (and accurate translation) but can lack equivalence of meaning. One simple (but expensive) way of dealing with this particular question would be to add an open-ended follow-up asking respondents to name those groups they were thinking of when answering the question.

Other solutions to this problem of contextualization have also been suggested. Przeworski & Teune (1967) recommend building scales composed of answers to several different questions to measure each concept. For each concept there needs to be a set of positively correlated questions that are valid in every country and can thus be used to measure the concept everywhere. They call these questions "identities." In addition to identities, some culturally specific items may be asked in a particular country, which are positively correlated with the identities and which can thus contribute to the scale for that country. Items like these are referred to as "equivalents." A good example of an equivalent for political participation might be voting in elections, which is impossible (and thus irrelevant) in nondemocratic countries but a good measure of participation in democratic countries. Thus, Przeworski & Teune recommend an "identity-equivalence" procedure for measurement, in which the same concept is measured with different scales in different countries, but each scale includes a common core of identities and possibly some culturally specific equivalents.

This kind of approach might be particularly useful if we are interested (as Przeworski & Teune are) in investigating relationships between variables in different countries. What we want, it could be argued, is the best measure of the concept in question for a particular country rather than a standard measure that is more appropriate in some countries than in others (and will therefore show weaker relationships in some countries than in others). The export of a standard measuring instrument, which was appropriate in the country of origin, may give an illusory appearance of comparability. A simple example is that of educational level. Number of years of completed education may well be an excellent

measure in some countries (e.g., the United States) where it is isomorphic with the educational system, but less appropriate in European systems with their selective educational systems. In Europe, respondents with a similar number of years of completed education may have had very different educational experiences, and measures of education that focus on qualifications achieved may be much more appropriate.

The difficulty with "functional equivalence," however, is the absence of any clear guidelines to indicate when equivalence has been achieved. These methods have not, therefore, acquired general support, nor are they useful for making population estimates.

King et al. (2004) have recently proposed a novel method of dealing with this problem using vignettes. Their starting point is Sen's (2002) observation that subjective assessments of health status may be highly context-dependent:

> The state of Kerala has the highest levels of literacy. . .and longevity. . .in India. But it also has, by a very wide margin, the highest rate of reported morbidity among all Indian states. . .. At the other extreme, states with low longevity, with woeful medical and educational facilities, such as Bihar, have the lowest rates of morbidity in India. . .. In disease by disease comparison, while Kerala has much higher reported morbidity than the rest of India, the United States has even higher rates for the same illnesses. If we insist on relying on self-reported morbidity as the measure, we would have to conclude that the United States is the least healthy in this comparison, followed by Kerala, with ill-provided Bihar enjoying the highest levels of health. In other words, the most common measure of the health of populations is negatively correlated with actual health. (Sen 2002, pp. 860–61)

As Sen explains, respondents in Bihar judge their health according to different yardsticks than do respondents in Kerala or the United States, thus leading to the paradoxical results.

King et al. (2004) attempt to build-in this subjectivity by establishing what responses probably mean in different contexts. Their approach is best illustrated by their main example. The ordinal question "How much say do you have in getting the government to address issues that interest you?" has been used to measure political efficacy in different countries. But it is not clear that similar answers mean the same thing in different political contexts. King et al. show that when the answers are taken at face value, political efficacy is greater in China (a country that has never had a democratic election) than in Mexico (a country that has had several elections, one of which recently removed a long-standing government). However, when respondents are asked to assess the level of political efficacy of characters in a set of fictional vignettes (see Table 3), it becomes clear that the Chinese are much more likely to ascribe greater political efficacy to the characters in the vignettes than are the Mexicans. Most strikingly, >40% of the Chinese respondents assessed their own level of political efficacy below that of the vignette character who "suffered in silence."

TABLE 3 Political efficacy vignettes (King et al. 2004)

1	[Alison] lacks clean drinking water. She and her neighbors are supporting an opposition candidate in the forthcoming elections that has promised to address the issue. It appears that so many people in her area feel the same way that the opposition candidate will defeat the incumbent representative.
2	[Imelda] lacks clean drinking water. She and her neighbors are drawing attention to the issue by collecting signatures on a petition. They plan to present the petition to each of the political parties before the upcoming election.
3	[Jane] lacks clean drinking water because the government is pursuing an industrial development plan. In the campaign for an upcoming election, an opposition party has promised to address the issue, but she feels it would be futile to vote for the opposition since the government is certain to win.
4	[Toshiro] lacks clean drinking water. There is a group of local leaders who could do something about the problem, but they have said that industrial development is the most important policy right now instead of clean water.
5	[Moses] lacks clean drinking water. He would like to change this, but he can't vote and feels that no one in the government cares about this issue. So he suffers in silence, hoping something will be done in the future.

King and his colleagues claim that vignettes can be used to "anchor" various survey questions, and they propose both a simple and a more sophisticated method for using the information from the vignettes to rescale self-assessment responses. For the simple method, they assume only (*a*) consistency of response over the vignettes and self-assessment question and (*b*) vignette equivalence across societies. This method has two particularly attractive aspects. First, it is not necessary to administer the vignettes in conjunction with the self-assessment question every time the question is asked; rather, the vignettes can be used to establish a rule for rescaling a particular question for a particular society, and that rule could be applied to earlier or later surveys (although the extent to which rescaling rules are stable over time has yet to be established). Second, it is not limited to tackling issues of comparability between societies; it could also be used to identify and tackle instances of questions and concepts being interpreted differently by different groups within the same society.

This method does seem to be one of the most promising for dealing with equivalence of meaning, but it is not yet clear whether appropriate vignettes can be developed to study other concepts of interest in public opinion research. To satisfy the assumption of vignette equivalence across societies, one needs to be able to construct a set of vignettes that are ranked in the same order in the different societies. This may prove to be a rather demanding requirement.

Cross-national differences in social desirability bias may also lead to problems for equivalence of meaning. For instance, Jones (1963) suggests that acquiescence bias (a tendency to agree with whatever proposition the interviewer poses) is greater in Southeast Asia than elsewhere. Acquiescence bias can often be dealt

with through good design, with measuring instruments that include questions with reverse wording. But this technique will not necessarily deal with social acceptability bias. If voting in elections is socially desirable, replacing the item "Do you agree that it is everyone's duty to vote?" with "Do you agree that it does not matter whether people turn out to vote or not?" is unlikely to get rid of social acceptability bias. Rather than treating social acceptability bias as a problem, we might regard it as a real difference between societies in their expression of public opinion. It would, however, be important to know if the relation between opinion and behavior varied systematically in more or less "courteous" societies. The inclusion of some behavioral items is therefore important for gaining a fuller understanding of public opinion.

The most severe difficulties in equivalence occur when the concept that a question is trying to tap into is less appropriate, or even nonexistent, in some contexts. Jowell (1998) gives the example of the difficulties that the cross-national ISSP team faced when developing their module on religion:

> [T]he Japanese delegation eventually came to the reluctant conclusion that there was no appropriate word or phrase in Japanese that approximated the concept of God. In the end, of course, they managed to come up with a doubtless somewhat tortuous circumlocution designed to get across the basic meaning of the Judaeo-Christian-Islamic concept of God. But beware of data that depend on such contrivances based on collegiality. (Jowell 1998, p. 172)

However, limiting a survey to concepts that can be measured in all countries could result in a set of anodyne questions that necessarily focus on common features of belief systems rather than on the distinctive features that have no cross-national equivalents. This point was made almost 40 years ago by Hodge and his colleagues in their classic comparative study of occupational prestige:

> It is quite possible that much genuine diversity in occupational-prestige systems not captured by our analysis is reflected in the relative placement of occupations that are not comparable across societies, or even across subsectors of any given society. In Tiryakian's study of the Philippines, for example, only a small number of occupations could be found about which it was sensible to ask both peasants in remote villages and the residents of Manila. (Hodge et al. 1966, p. 311)

The globalization of public opinion research does seem, therefore, to have led to the use of standardized closed questions around the world—something of a standard global product. The strong assumptions underlying this product are well known, but the alternatives, such as the use of supplementary open-ended questions or the vignette method, are likely to be time-consuming and expensive. A program of methodological work to determine which measuring instruments are particularly prone to problems of "equivalence of meaning" appears overdue.

THE INTELLECTUAL ASSUMPTIONS OF CROSS-NATIONAL SURVEYS

Rather broader intellectual questions should also be asked about the globalization of public opinion research. It could be argued that standardized questions are not the West's only global export; there is also (perhaps in consequence) the export of a broader intellectual understanding of public opinion and of the appropriate concepts for describing public opinion. Particular questions also embody particular intellectual traditions.

This can perhaps be seen most clearly in the 1960s export of the Michigan model of election studies, with its sociopsychological intellectual framework, rather than the Columbia model, with its more sociological framework. Currently dominant intellectual approaches for understanding electoral behavior tend to be individualistic, often based on rational choice theory. The questions asked, for example, in the CSES module cover the kinds of concepts (such as attitudes toward leaders, the left-right domain, and judgments of economic performance) that have been associated with this type of intellectual framework.

We do not mean to suggest that attitudes toward leaders or the left-right dimension are unimportant aspects of public opinion; however, we can question whether the intellectual framework implicit in this particular battery of questions is as appropriate for, say, India (not actually a member of CSES) as it is for the United States or Britain. An emphasis on attitudes toward issues or the national economy may well be appropriate in countries with literate electorates who read newspapers or view televised debates about key issues, but in other parts of the world—such as India—the group processes of the sort investigated originally by the Columbia school [and more recently by Huckfeldt (1984) and Huckfeldt & Sprague (1995)] might be equally valuable in understanding public opinion.

Nor do we wish to be critical of the Michigan model and its export, which was at the time a highly progressive development. However, since many cross-national series (rightly) aim to chart change over time, there is a tendency to maintain previous questions. This has perhaps led to an institutional conservatism in questionnaire content and, hence, in intellectual approach.

A second, related concern is the notion of public opinion itself. In the United States, Converse (1964), one of the key members of the Michigan team, raised the fundamental issue of attitudes and nonattitudes. It is beyond the scope of this review to discuss the enormous debate that ensued, but it may be of particular contemporary relevance with the globalization of attitude research. There is clearly a risk that the administration of standard questions in surveys across the world might give a misleading impression that publics in different countries do vary in their attitudes toward the topics posed. The various barometers, for example, will often tabulate mean scores toward a topic, e.g., international trade relations, by country. But is there a danger that these findings are in part an artifact of the survey method? Might it not be the case that in some countries there is no public debate

about international trade relations, and thus no good grounds for supposing public opinion is more or less supportive of this issue than in other countries?

An empirical example of this kind of problem can be seen with the measurement of economic left-right values. We can ask the same battery of questions that tap this dimension across the world (and there are questions in the WVS that can be used for this purpose) and then calculate mean scores on a scale derived from this battery. However, the more instructive finding may not be the calculated mean scores but rather the finding that the internal reliability of the scale is almost meaninglessly low in some countries. Tilley (2002), for example, shows that a scale of economic left-right values makes much better sense (as judged by internal reliability) in Northwestern Europe than it does in Latin America or Eastern Europe. This suggests that in these latter areas, either people do not respond to different economic issues in a way that is consistent with a unifying ideological position, as in Northwestern Europe, or there are problems with the applicability of the questions in the different contexts there.

This is not, in itself, a criticism or limitation of cross-national survey research, as checking internal reliability of scales is routine. Indeed, one important task of cross-national research could be to detect how far there actually is public opinion on certain domains.

But even if the problem of nonattitudes has been addressed, a national mean score may not be appropriate in a divided society. We referred above to the problem of divided societies such as Northern Ireland in the context of fixing sample sizes. In such cases, it is quite conceivable that no one in the society actually holds the national mean-score opinion, because public opinion is polarized rather than normally distributed. This calls into question the idea of a single measure of national public opinion and suggests that the relevant unit of analysis may not be the state but some subunit. This may be particularly important in multination states, such as the many countries in which nations and states are truly misaligned. Because the state is the usual administrative unit (both for sampling frames and for political decisions), it is understandable that public opinion research has tended to take the state, rather than the nation, as the unit of analysis. The subjective quality of "the nation" makes it perhaps more appropriate as the subject matter of public opinion research than as the basis of its conduct. However, there are many cases, such as Canada, Belgium, and several Indian states, where territorial subunits of the overall state make both effective administrative units and meaningful linguistic or ethnic groupings.

CONCLUSIONS

As the overview of cross-national survey programs presented at the beginning of this paper showed, globalization of academic public opinion research has indeed been occurring—but it is still far from complete. Although the number of countries included in the major cross-national programs continues to grow, there is still

a predominance of the core established democracies of North America, Western Europe, and Australia. In addition, globalization of public opinion surveys has not entailed a straightforward spread of a standardized "product" throughout the world in terms of survey conduct. Some key features—systematic sampling and standardized questionnaires administered by trained interviewers—have been present in virtually all the surveys included in the various cross-national programs, but sample designs and fieldwork methods remain highly diverse. To a degree, this reflects the diversity of survey research methods in the core countries, but inevitably it also reflects the varying conditions under which survey research has to be conducted in different parts of the world. These differences can give rise to large, but often unknown, differences in quality between surveys and between countries. The World Association for Public Opinion Research has suggested rules of practice regarding the documentation of surveys (http://www.unl.edu/WAPOR/ethics.html), but these rules are all too rarely followed.

Globalization has tended to lead to a more standardized product with respect to questionnaire content, but this may itself be another source of noncomparability if the same questions are interpreted differently in different contexts. There has also perhaps been a globalization of intellectual assumptions and frameworks, which partly reflects the assumptions and theories of the questionnaires' originators. Intellectual frameworks developed to understand public opinion in urbanized, literate, and relatively homogeneous societies should not be assumed to be appropriate for other contexts.

A number of writers have suggested rules for the conduct of cross-national research designed to address some technical aspects of noncomparability (e.g., Scheuch 1968; Kuechler 1987, 1998; Jowell 1998), and the ESS is a particularly impressive example of the implementation of rules designed to secure high-quality surveys in all participating countries. Other scholars, however, suggest that if a country is of particular theoretical interest, even low-quality data might be acceptable if nothing better is available (e.g., Inglehart 1997, p. 347).

This debate is, in some ways, akin to debates in evidence-based medicine. One could insist on the "gold standard"—in the medical case, double-blind placebo-controlled experiments. However, because experiments of this sort are often hard to come by, current medical practice is to consider results from other, less rigorous research designs as well when deciding on treatment, although these are given less weight. Grading systems have been developed to aid in this delineation. Table 4 shows one such grading system (Eccles et al. 1998). (For more details on evidence-based medicine, see http://www.cochrane.dk/cochrane/handbook/hbook.htm.)

Our own sympathies are with the evidence-based medicine approach. We agree with the medical guidelines that the most weight should be given to meta-analyses of high-quality data and that the least weight should be given to nonsystematic research. It may be more difficult than in the medical case to achieve consensus on the intermediate grades, but for the sake of argument, we propose a scheme along the lines of Table 5. We suspect that the great majority of the surveys involved in the cross-national programs described in this paper fall into the proposed categories II

TABLE 4 Categories of evidence about treatment (Eccles et al. 1998)

Ia	Evidence from meta-analysis of randomized controlled trials
Ib	Evidence from at least one randomized controlled trial
IIa	Evidence from at least one controlled study without randomization
IIb	Evidence from at least one other type of quasi-experimental study
III	Evidence from nonexperimental descriptive studies, such as comparative studies, correlation studies, and case-control studies
IV	Evidence from expert committee reports or opinions and/or clinical experience of respected authorities

and III. Although we sympathize with the aim of the ESS to drive up standards so that they all come into our top category, we believe the reality is that most surveys will continue to be in category II at best, and that public opinion analysts need to adjust to this reality. We consider it a priority, therefore, for methodologists to establish empirically whether findings from random-route samples, for example, show different patterns from those obtained by strict probability samples with high response rates. (In the evidence-based medicine field, methodologists have been at pains to emphasize that scales of data quality have to be assessed empirically in just the same way as other measuring instruments, and our proposed grading ought also to be validated by empirical research.)

The need for schemes of this kind arose in medicine because of the sheer volume of research, of highly uneven quality, which the individual practitioner could not hope to study in the required depth. Because public opinion research does not have the same volume of studies as medicine, the need for meta-analyses may not be so evident. However, as Table 1 showed, many countries now participate in several different cross-national programs, and so the possibility of meta-analysis is there. Moreover, the unevenness of quality is certainly an issue, and where studies fall short of the gold standard (as the great majority do), it becomes all the

TABLE 5 Proposed weighting of sample surveys

Ia	Evidence from meta-analysis of systematic samples with full documentation
Ib	Evidence from at least one probability sample with response rate of 70% or above and full documentation following WAPOR[a] rules of practice
II	Evidence from at least one systematic representative survey (including random route or quota methods or probability samples with response rate below 70%) with full documentation following WAPOR rules of practice
III	Evidence from at least one systematic survey lacking full documentation
IV	Evidence from nonsystematic nonrepresentative studies, such as "snowball" samples

[a]WAPOR, World Association for Public Opinion Research.

more important to compare results from several different studies rather than to rely solely on the one that happens to have participated in the particular cross-national program. In short, data of lower than ideal quality should be checked against the results of other surveys that cover the same topics.

To be sure, one reason why meta-analysis is rarely carried out is that the very standardization of questions within cross-national programs tends to lock the user into one particular program, making it difficult or impossible to compare findings with those of other programs. Each program tends to have developed its own, idiosyncratic questions and question wording even when it is studying a relatively common concept, such as left-right values. Again, public opinion researchers may have something to learn from medicine, where there tend to be many standardized (and validated) measuring instruments that can be used in any study on the topic. Inglehart's postmaterialism index is one of the few such instruments in public opinion research. King's political efficacy vignettes deserve to become another. Although there have been many criticisms of the postmaterialism index, the advantage of such a widely used index is that its properties are now well known. The wider use of such instruments in public opinion research would certainly aid meta-analysis.

We have raised several serious caveats about the use of standardized measures in different cultural contexts, and these caveats would necessarily apply to the use of standard instruments such as the postmaterialism index as well as to the standardized questions favored by cross-national programs. (However, we must emphasize that our main reason for advocating standard instruments is to permit cross-checking of results *within* a given country, not necessarily to aid cross-national comparison.) Given the current state of uncertainty about the equivalence of standard items across cultural contexts, we advocate the use of a range of methods and the evaluation of these methods by systematic methodological research. We have, for example, advocated the use of supplementary open-ended material [as recommended by Schumann & Presser (1981) in their classic work on questions and answers] and the use of King's vignette method. Przeworski & Teune's method of identities and equivalences also deserves to be revived; we believe that, in addition to the common questions designed by the central coordinating committee of the national program, individual participating countries should be encouraged to include their own questions that are faithful to the local context.

Comparisons of public opinion between countries at a single point in time are the most vulnerable to problems of survey quality and comparability, even if their findings are statistically significant according to standard tests. The claim that Country X has greater support for democracy than Country Y could be misleading for numerous reasons; by contrast, a claim (supported by significance tests) that over the past five years support for democracy has increased in Country X but declined in Country Y is likely to be more robust. If there is continuity of methodology, the analysis of change in public opinion over time within countries and the comparison of trends across countries are much safer than single-time-point

comparisons, since context-specific interpretations of questions and idiosyncrasies in sample design are likely to be relatively stable and should effectively cancel out when the difference between two time points is calculated within a country.

Repeated cross sections are valuable (though vulnerable to differences in survey quality over time given declining response rates), but change is best calculated with a panel study in which the same individuals are interviewed more than once. It is unfortunate that no cross-national panel studies carry a substantial number of social attitude items of interest to political scientists. Although we began our discussion by presenting the GSS as our gold standard in survey research, there is a growing tendency in social science to see the cross-section survey as second-best to a well-conducted panel. Perhaps a new process of the globalization of panel studies ought to begin.

Finally, cross-national survey researchers need to pay more attention to the unit of analysis. The state is the usual unit, and there are some good, practical reasons for this. But the nature of states is often contested by independence movements, and it is not always clear that the state is sociologically the most appropriate unit, especially in diverse or divided multination states. To take an extreme example, Czechoslovakia was a single state at the time of the 1991 WVS and was covered with a single survey, but now there would be separate surveys in the independent states of the Czech Republic and Slovakia. However, the "velvet divorce" between the two parts of Czechoslovakia was presumably related to cultural and social differences of exactly the sort that public opinion research addresses. Similarly, Quebec almost certainly warrants a separate survey from the rest of Canada, Flanders from Wallonia, and Scotland from England. To accept the state as the unit of analysis is to accept a questionable normative position.

We recognize that many of our suggestions for larger samples, panel studies, vignettes, and open-ended and country-specific questions are expensive. However, the Afrobarometer and the new State of Democracy in South Asia program indicate that there are practicable alternative ways forward. Rather than a simple export of Western methods, assumptions, and intellectual frameworks to nonwestern societies, public opinion research might benefit from imports in the reverse direction.

APPENDIX: SURVEY-RELATED WEB SITES

Afrobarometer: http://www.afrobarometer.org/

- Sampling Protocol: http://www.afrobarometer.org/SamplingProtocol.pdf

American National Election Studies: http://www.umich.edu/~nes/

British Election Studies (University of Essex): http://www.essex.ac.uk/bes/index.html/

Candidate Countries Eurobarometer: http://europa.eu.int/comm/public_opinion/

- Data Archive: http://www.gesis.org/en/data_service/eurobarometer/cceb/index.htm

Comparative Study of Electoral Systems: http://www.umich.edu/~cses/

East Asia Barometer. http://www.eastasiabarometer.org/

Eurobarometer. http://europa.eu.int/comm/public_opinion/

- Data Archive: http://www.gesis.org/en/data_service/eurobarometer/

European Social Survey: http://www.europeansocialsurvey.org/

European Values Survey: http://www.europeanvalues.nl/

Gallup Voice of the People Survey: http://www.voice-of-the-people.net/

General Social Survey: http://www.norc.uchicago.edu/projects/gensoc.asp/

Global Barometer: http://www.globalbarometer.org/

International Social Survey Programme: http://www.issp.org/

Latinobarómetro: http://www.latinobarometro.org/

New Europe Barometer: http://www.cspp.strath.ac.uk/

Pew Global Attitudes Project: http://people-press.org/pgap/

State of Democracy in South Asia: http://www.lokniti.org/projects.htm#sdsa

World Association of Public Opinion Research: http://www.unl.edu/WAPOR/

- Code of Professional Ethics and Practices: http://www.unl.edu/WAPOR/ethics.html

World Values Survey: http://www.worldvaluessurvey.org/

The *Annual Review of Political Science* is online at http://polisci.annualreviews.org

LITERATURE CITED

Almond GA, Verba S. 1963. *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton, NJ: Princeton Univ. Press

Almond GA, Verba S. 1989. *The Civic Culture Revisited*. Newbury Park, CA: Sage

Barnes SH, Kaase M, Allerbeck KR, Farah BG, Heunks F, et al. 1979. *Political Action*. Beverly Hills, CA: Sage

Lazarsfeld P, Berelson B, Gaudet H. 1948. *The People's Choice*. New York: Columbia Univ. Press

Berelson BR, Lazarsfeld PF, McPhee WN. 1954. *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago: Univ. Chicago Press

Bulmer M. 1998. The problem of exporting social survey research. *Am. Behav. Sci.* 42:153–67

Butler D, Stokes D. 1974. *Political Change in Britain*. London: Macmillan

Campbell A, Gurin G, Miller WE. 1954. *The Voter Decides*. Evanston, IL: Row, Peterson

Campbell A, Converse PE, Miller WE, Stokes DE. 1960. *The American Voter*. Chicago: Univ. Chicago Press

Colley L. 1992. *Britons: Forging the Nation: 1707–1827*. New Haven, CT: Yale Univ. Press

Converse PE. 1964. The nature of belief systems in mass publics. In *Ideology and*

Discontent, ed. DE Apter, pp. 206–61. New York: Free

Converse PE. 1987. *Survey Research in the United States: Roots and Emergence 1890–1960*. Berkeley: Univ. Calif. Press

Crespi I. 1989. Public opinion. In *International Encyclopedia of Communications*, ed. E Barnouw, 3:386–91. New York: Oxford Univ. Press

Eccles M, Freemantle N, Mason J 1998. North of England evidence based guidelines development project: methods of developing guidelines for efficient drug use in primary care. *Br. Med. J.* 316:1232–35

Eldersveld SJ, Ahmed B. 1978. *Citizens and Politics: Mass Political Behaviour in India*. Chicago: Univ. Chicago Press

Gallup G. 1948. *A Guide to Public Opinion Polls*. Princeton, NJ: Princeton Univ. Press

Groves RM. 1987. Research on survey data quality. *Public Opin. Q.* 51:S156–72

Groves RM. 1989. *Survey Errors and Survey Costs*. New York: Wiley

Halman L. 2001. The European Values Study: a third wave. *Source Book of the 1999/2000 European Values Study Surveys*. Tilburg, The Netherlands: EVS, WORC, Tilburg Univ.

Harkness J. 1999. In pursuit of quality: issues for cross-national survey research. *Int. J. Soc. Res. Methodol.* 2:S125–40

Harkness J, Schoua-Glusberg A. 1998. Questionnaires in translation. In *Cross-Cultural Survey Equivalence*. Mannheim, Ger.: Zentrum für Umfragen, Methoden und Analysen

Hodge RW, Treiman DJ, Rossi PH. 1966. A comparative study of occupational prestige. In *Class, Status, and Power: Social Stratification in Comparative Perspective*, ed. R Bendix, SM Lipset, pp. 309–21. New York: Free. 2nd ed.

Hovland CI, Lumsdaine AA, Sheffield FD. 1950. *The American Soldier, Volume III: Experiments in Mass Communication*. New York: Wiley

Huckfeldt RR. 1984. Political loyalties and social class ties: the mechanisms of contextual influence. *Am. J. Polit. Sci.* 78:399–417

Huckfeldt RR, Sprague J. 1995. *Citizens, Politics and Social Communication: Information and Influence in an Election Campaign*. Cambridge, UK: Cambridge Univ. Press

Inglehart R. 1990. *Culture Shift in Advanced Industrial Society*. Princeton, NJ: Princeton Univ. Press

Inglehart R. 1997. *Modernization and Postmodernization: Cultural, Economic and Political Change in 43 Societies*. Princeton, NJ: Princeton Univ. Press

Jones EL. 1963. The courtesy bias in Southeast Asian surveys. *Int. Soc. Sci. J.* 15:70–76

Jowell R. 1998. How comparative is comparative research? *Am. Behav. Sci.* 42:168–77

King G, Murray CJL, Salomon JA, Tandon A. 2004. Enhancing the validity and cross-cultural comparability of measurement in survey research. *Am. Polit. Sci. Rev.* 94:191–207

Kish L. 1965. *Survey Sampling*. New York: Wiley

Kish L. 1994. Multipopulation survey designs: five types with seven shared aspects. *Int. Stat. Rev.* 62:167–86

Klein S, Harkness J. 2003. *ISSP Study Monitoring 2001: Report to the ISSP General Assembly on Monitoring Work Undertaken for the ISSP by ZUMA, Germany*. Mannheim, Ger.: Zentrum für Umfragen, Methoden und Analysen

Kuechler M. 1987. The utility of surveys for cross-national research. *Soc. Sci. Res.* 16: 229–44

Kuechler M. 1998. The survey method: an indispensable tool for social science research everywhere? *Am. Behav. Sci.* 42:178–200

Lyberg L, Biemer P, Collins M, de Leeuw E, Dippo C, et al. 1997. *Survey Measurement and Process Quality*. New York: Wiley

Lynn P. 2003a. Developing quality standards for cross-national survey research: five approaches. *Int. J. Soc. Res. Methodol.* 6:323–36

Lynn P. 2003b. *Development of a sampling method for household surveys in post-war Bosnia and Herzegovina*. Work. Pap., Inst. Soc. Econ. Res., Univ. Essex, Colchester, UK

Lynn P, Hader S, Gabler S, Laaksonen S. 2004. *Methods for achieving equivalence of samples in cross-national surveys: the European Social Survey experience*. Work. Pap., Inst. Soc. Econ. Res., Univ. Essex, Colchester, UK

Marsh C. 1982. *The Survey Method*. London: Allen and Unwin

Marshall G. 1998. *Oxford Dictionary of Sociology*. Oxford, UK: Oxford Univ. Press

Norris P. 2004. From the civic culture to the Afrobarometer. *APSA-CP Newsl.* 15:2

Otava J. 1988. Public opinion research in Czechoslovakia. *Soc. Res.* 55:247–60

Park A, Jowell R. 1997. *Consistencies and Differences in a Cross-National Survey*. London: Social and Community Planning Research

Przeworski A, Teune H. 1967. Equivalence in cross-national research. *Public Opin. Q.* 30:551–68

Rohme N. 1997. *The Freedom to Publish Opinion Polls: Report on a Worldwide Study*. Amsterdam: ESOMAR

Rose N, Osborne T. 1999. Do the social sciences create phenomena? The case of public opinion research. *Br. J. Sociol.* 50:367–96

Rossi P. 1964. Four landmarks in voting research. In *Readings in Political Parties and Pressure Groups*, ed. F Munger, D Price, pp. 304–47. New York: Thomas Y. Crowell

Scheuch EK. 1968. The cross-cultural use of sample surveys: problem of comparability. In *Comparative Research Across Cultures and Nations*, ed. S Rokkan, pp. 176–209. Paris: Mouton

Schumann H, Presser S. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic

Sen A. 2002. Health: perception versus observation. *Br. Med. J.* 324:860–61

Sinnott R. 1998. Party attachment in Europe: methodological critique and substantive implications. *Br. J. Polit. Sci.* 28:627–50

Smith TW. 2004. Freedom to conduct public opinion polls around the world. *Int. J. Public Opin. Res.* 16:215–23

Spangenberg F. 2003. *The Freedom to Publish Opinion Polls: Report on a Worldwide Update*. Amsterdam: ESOMAR

Stouffer SA, Suchman EA, DeVinney LC, Star SA, Williams RM Jr. 1949a. *The American Soldier, Volume I: Adjustment During Army Life*. New York: Wiley

Stouffer SA, Lumsdaine AA, Lumsdaine MH, Williams RM J Jr, Smith MB, et al. 1949b. *The American Soldier, Volume II: Combat and Its Aftermath*. New York: Wiley

Stouffer SA, Guttman L, Suchman EA, Lazarsfeld PF, Star S, Clausen JA. 1950. *The American Soldier, Volume IV: Measurement and Prediction*. New York: Wiley

Tilley J. 2002. Is youth a better predictor of sociopolitical values than is nationality? *Ann. Am. Acad. Polit. Soc. Sci.* 580:226–56

Warwick DP, Osherson S. 1973. Comparative analysis in the social sciences. In *Comparative Research Methods: An Overview*, ed. DP Warwick, S Osherson, pp. 3–41. Englewood Cliffs, NJ: Prentice-Hall

Worcester RM. 1987. The internationalization of public opinion research. *Public Opin. Q.* 51:S79–85

Annual Review of Political Science
Volume 8, 2005

CONTENTS