

The Ups and Downs of Cross-National Survey Research

Tom W. Smith

NORC
University of Chicago

GSS Cross-National Report No. 8

May, 1988
Revised December, 1988

This research was done for the General Social Survey project directed by James A. Davis and Tom W. Smith. The project is funded by the National Science Foundation, Grant No. SES-8747227.

What is important in comparative research is the exchange of findings-replicative testing of the same theories in varying social contexts. When the findings are similar, evidence accumulates to support their generality.... When findings are different, we need to explain these differences.... What we need to do is control for theoretically significant system-level differences that can be expressed as variables (Przeworski and Teune, 1971, p. 134).

In the social sciences' search for lawfulness, cross-national survey research is one of the most rewarding and punishing of pathways. When the same models work across nations with the same hypotheses confirmed, the generalizability of our findings are greatly strengthened and our understanding of human society advanced. When the same models fail to work across nations and the same hypotheses are not supported, we know that some additional variables must be incorporated into our models. This search for the right additional variables is often arduous and, as we shall see, may lead us through methodological bogs, but when the proper specification or the missing variables can be discovered our understanding of human society is greatly expanded (Kohn, 1987; Rokkan, Viet, Verba, and Almsay, 1969; Niessen and Peschar, 1982; Armer and Marsh, 1982; Szalai and Petrella, 1977).

Basic Approaches

When confronted with a cross-national difference between two or more nations, there are two basic approaches towards explaining the difference, the idiographic and continuum (Jasper, 1987). The idiographic approach looks to unique, special case explanations for the difference. Usually some particular

historical event or distinctive cultural trait is offered to explain the variation. This approach is qualitative and tends to be used in the disciplines of history, anthropology, and personality psychology. It is also more common when only a small number of nations are being compared and especially when only two are under investigation. An example of this approach is the common "except for the South" caveat that is often used to describe nineteenth century America, such as "America was a nation of small, family farms, except for the South," "America had a two party system, except for the South," or "Americans believed that 'all men are created equal,' except for the South." In each of these cases the reason for the Southern exception was slavery, which, as befits an idiographic perspective, was widely known as that "peculiar institution."

On the other hand, the continuum approach assumes that nations vary along various underlying scales and that differences between nations arise from their different values on these variable scales. This approach is quantitative and multivariate and is most frequently employed in economics, sociology, and political science. It tends to be used when a large number of nations are being compared. An example would be a study that explained differences in life expectancy across nations by the number of doctors and hospital beds per capita and which in turn explained disparities in these medical resources by economic level (e.g. per capita GNP or energy consumption).

Of course these are not mutually exclusive techniques. A particular scholar's analysis could blend together these two approaches or use them in combination. For example, a detailed historical and cultural understanding of the nations involved might well suggest variables that could be quantified and then utilized in a multivariate analysis.

Alternatively, a residual or outlier analysis might indicate that certain variables explained most of the cross-national variation, but that one nation or perhaps a group of nations related in some way (e.g. by a common heritage or geographic proximity) deviated from the generally good fit of the model. One might have to offer idiographic explanations for such outliers. This residual approach has great promise since it allows us to 1) develop a model that better explains the cross-national variation and 2) fruitfully combine qualitative and quantitative techniques (and disciplines such as history and economics).¹ The dangers are that 1) we might over use particularistic explanations for outliers so that we end up accepting as correct a basic model that really is misspecified and 2) we accept unique, historical explanations that by their

¹ For example, Treiman (1977) found that occupational prestige was determined by similar factors across almost all societies. He did find certain outliers such as the higher prestige of miners in Communist countries and developed historical and structural explanations for this pattern.

nature cannot be subjected to quantitative verification instead of uncovering more general explanations that both could apply in more circumstances and would be subject to empirical verification. For example, if we were conducting a study of tolerance of cultural pluralism and found that Canada was more tolerant than the general model predicted, we could treat Canada as a special case (or as a dummy variable in the multivariate analysis) and make reference to its dual colonial history and two charter groups. Rather than going to this particularistic (and unconfirmable) approach, we might decide that Canada was more tolerant because of the size of its minority language population. We could then add a variable, % minority language speakers, that might explain why Canada was an outlier as well as improve the fit of the model across all countries. Of course it might not be readily possible to come up with general variables to explain outliers. There may be complicated, unique historical traditions in Canada that truly explain its special leanings which could not be readily reduced to quantifiable, general variables that could be coded for all nations.

Measurement Difficulties

Whether an idiographic, continuum, or combination approach is utilized, cross-national survey research offers great potential for both theoretical refinement of our understanding of human society and a means to test our theories. But cross-national research must also overcome great barriers to achieve its potential. The barriers to the successful completion of cross-

national survey research (and many other related types of cross-national research) are so great that a study of articles in Comparative Politics and Comparative Political Studies from 1968 to 1982 found that 62% actually analyzed only one country (Jackman, 1985).

First, there are the organizational difficulties. The administration of the cross-national survey efforts is always complex, involving the coordination of principal investigators, funding sources, data collection and research institutions, and perhaps governments. Next, the cost is high. Roughly speaking the cost is a multiple of the number of nations participating. The more the number of participating nations, the greater the intellectual potential, but the higher the cost. In addition, the planning, execution, and analysis of the research design takes much longer than any single nation effort. From my experience with the International Social Survey Program (ISSP) (CROSS REFERENCE TO OTHER ARTICLE; Smith, 1986; Kuechler, 1987), I would say that it takes 3-4 times as long to do a study from start to finish across a half dozen nations as it does for a similar study in one nation.

Second, there are basic measurement issues that make cross-national survey research extremely challenging (Dogan and Pelassy, 1984; Rokkan, Viet, Verba and Almsay, 1969; Vallier, 1971). As I once remarked, "Equivalency, the design of survey

instruments that are efficient, reliable, and valid not only in a single society, but across several nations, is a difficult task." Most broadly, cultural differences between nations may hinder equivalency. Presumably surveys of dissatisfaction with government would get meaningless responses in repressive regimes that suppressed all dissent. Or surveys of women might not be readily possible in fundamentalist Islamic nations, or at least not if using home interviews by male interviewers. Or certain institutions may vary across countries, making comparisons difficult. For example, in the United States the President is both head of government and head of state, while in most European governments these offices are held by different people (e.g. in England respectively by the prime minister and the monarch). Questions about Reagan and Thatcher usually ignore these differences in their institutional roles.²

Of course the cultural difference that most often creates problems is language. There are well-established procedures for parallel and back translation to insure an optimum linguistic match and these must naturally be rigorously employed. These techniques do not insure that true equivalency is achieved in survey questions however. Even given the most careful of translations it is nearly pointless to compare any two questions

² In fact studies of political behavior and state-citizen relations are among the most common type of cross-national survey research (Almond and Verba, 1963; 1980; Barnes, et al., 1979; Inglehart, 1972; Verba, Nie and Kim, 1971; 1978).

that employ abstract concepts and subjective response categories. While it is probably possible to ask effectively exact equivalents to the questions "In what year were you born?" or "Did you vote in the last national election?" It is highly doubtful that the responses to the query "Are you very happy, pretty happy, or not too happy?" are comparable. In all likelihood the closest linguistic equivalent to "happy" will differ from the English concept in various ways, perhaps conveying different connotations and tapping other dimensions (e.g. satisfaction), but at the minimum probably expressing a different level of intensity (say on an absolute bliss to sadness continuum). Similarly the adjectives "very", "pretty", and "not too" are unlikely to have precise equivalents. On an absolute intensity scale running from 0 to 10, "very" might rank an 8 and pretty as a 5 while the closest counterparts in a second language might be at 7.2 and 5.8 or 8.4 and 4.8, a difference that would certainly either produce different marginals or, perhaps, similar marginals that disguised differences in the absolute happiness distributions.³

Methodological Solutions

At least four solutions have been offered and each is certainly worth exploring. The first solution is to use numerical scales (e.g. ratio-level, magnitude measurement scales or thermometers). Numerical scales are believed to reduce

³ For an example on the numerical evaluation of adjectives and other modifiers see Hans-J. Hippler and Norbert Schwarz, 1986.

problems by providing a universally understood set of categories that have precise and similar meanings (e.g. 1,2,3 or 2:1, 3:1 have cross-linguistic equivalents) and that there is no need to come up with language labels to try and denote the intensity of each category. The problems are 1) that many (but not all) of these scales are more complex than simpler, verbal items and that this may actually increase the non-comparability problem, 2) that such an approach does not address the problems of variation in the meaning and strength of the basic abstraction involved, and 3) that different cultures may vary in their understanding of and ways of responding to numerical scales.

Another possibility, in a sense the opposite of the numerical approach, is the "keep-it-simple-stupid" approach. For surveys this would mean to use only dichotomies. Presumably (but it could certainly use empirical verification) yes/no, favor/oppose, etc. have similar cutting points across languages (or at least that there are equivalent pairs, even if these might not be the optimum examples). The argument is that it may be difficult to determine, because of language, where someone sits on a continuum, but comparatively easy (in the aggregate) to determine on which side of a tipping point people are. But this approach also begs the question whether the underlying concept is being understood in a similar enough fashion in both languages and results in less information than a question with say seven categories. At least in terms of crude number of response

categories one would have to ask three to six dichotomies (depending on the handling of DKs) to get as many response categories distinguished as one seven point question.

The third approach is to attempt to calibrate the scales by determining the strength of the verbal labels used. This would permit the adoption of labels for the happiness question that cut the same points of the underlying bliss-sadness continuum. Similarly by both linguistic analysis and numerical evaluation it would be possible to determine whether happiness was an equivalent stimulus in two languages/cultures. Ideally, it might be possible to determine that certain verbalizations and scales were similar between languages and that such equivalency held across their use in diverse questions. As such it might be possible to develop a standard set of response categories that could be used across many different questions. This approach necessitates extensive prior work of the meaning of words and the strength of adjective labels. It also typically assumes that modifiers will have similar strength across a wide range of applications (i.e. in all questions).

The final approach is to adopt multiple indicators. This approach differs in one significant way from the typical psychometric scale approach in that not only would multiple questions be employed, but different response scales would be used. Consider the following scheme to assess whether the French

or English are higher on psychological well-being.

- A. General happiness
- B. Overall satisfaction
- C. Domain-specific satisfaction
- D. Bradburn's affect-balance scale

While a Anglo-French comparison on any one of these question scales would be suspect because of the language ambiguities outlines above, it is possible to get unambiguous results from the above if the results are consistent (i.e. the French leading or trailing the British on all measures). Of course, if all the results are consistent then it was not necessary (except for the great gain in precision and richness) to have used more than one indicator. However, the difference is that with one indicator you never know whether the observed differences are social or only linguistic. What is needed is at least three indicators of the same construct. If one has two and they disagree it might be that one is "real" and the other only linguistic. With three you have a "tie breaker" (Think how this phrase might translate into another language. It might come out as an abuser of neckties or an untier of knots. Of course good translation and especially back translation will avoid these pitfalls.) in which you can hopefully isolate the odd-linguistic case from the two consistent social differences. Of course life need not be so simple. Few results would be purely "linguistic" or "social". Questions may differ because they fail to tap the same dimension or may be tapping a single construct in one language, but two distinct

constructs in a second language, or because of other problems. However we feel that three carefully translated questions, pretested as reliable in each individual language, and guided by clear and reasonably sound social science theory should allow us to avoid most of the larger problems, most of the time (but not all problems all of the time).

Of course these techniques are not mutually exclusive, but rather directly complementary. Adopting multiple indicators is the first step, but in the design of these multiple indicators it would be quite sensible to use a numerical scale for some indicators and of course extremely useful to learn more about differences in concept words and the intensity of modifiers via open-minded questions and non-survey textual analysis in the first case and via scaling and relating techniques in the latter case.

Even given the best precautions, measurement artifacts can slip thorough. As a result we must be prepared to detect such problems at the analysis stage. Basically what we can do is inspect the data for anomalies such as high levels of Don't Knows which may indicate a lack of understanding or relevance in a particular country and unusual marginals or relationships which may indicate not true differences, but only measurement artifacts. When suspicious signs are found, we must review the questions wordings, data codings, etc. to see if there are any

signs of unintended differences. For example, in the 1987 ISSP survey on social inequality one item asked people to rate their social standing by placing themselves on a 10-point ladder. The Dutch data showed many more people in the bottom three rungs than any other nation. Inspection of their instrument indicated no apparent problem with the wording, but their ladder widened at the bottom, obviously suggesting to respondents that more people were in the lower rungs than the middle and top (Smith, 1988). While one will occasionally find such a smoking gun, often no obvious problems will be discovered.

That leaves the investigator to ponder, is it real or is it artifact? Assuming that neither the triangulation procedure described above, the methods review, nor any other internal analysis clarifies the issue, one is left to apply certain external tests. Perhaps first and foremost is to ask whether the difference confirms to our a priori theory. Second, if no prior theory had been formed about the particular difference observed, one might examine other research including histories and single nation studies to see if the difference might be consistent with known attributes of the country. Third, in special cases one can eliminate language as a probable factor if the difference appears in some countries using the same language and not in others. For example, the US and to a lesser extent Australia are less supportive of the welfare state than are Western European nations. Since Great Britain is English-speaking, but resembles

its continental brothers more than its colonial children, we can safely conclude that language is not the explanation for the cross-national differences (Smith, 1987; 1988). In another special case we can look at multi-lingual countries (e.g. Canada, Switzerland, Belgium) to see if a national difference is really due to country and not language (Inglehart and Rabier, 1985). For example, if the French speaking Walloons in Belgium are less happy than the French, while the Dutch speaking Flems are less happy than people in the Netherlands, then we can be reasonably sure that there is something about being Belgian that makes people less happy than being French or Dutch. Of course the pattern need not always be so clear cut. Finally, in some circumstances the best thing may be to try to replicate the result, perhaps by developing alternative indicators that should show the same difference if the pattern is real rather than artifactual.

Summary

Overall, the "downs" of cross-national survey research are neither trivial nor intractable. They are inherent and unavoidable problems that must and can be copted with. They can be minimized, but not eliminated. Despite these problems, the "ups" of cross-national survey research are well-worth the effort. Profound insights into the human condition are more likely to emerge from such comparative perspectives than from alternative approaches.

References

- Almond, Gabriel A. and Verba, Sidney, The Civic Culture: Political Attitudes and Democracy in Five Nations. Princeton: Princeton University Press, 1963.
- Almond, Gabriel A. and Verba, Sidney, The Civic Culture Revisited. Boston: Little, Brown, 1980.
- Armer, M.J. and Marsh, R. M., Comparative Sociological Research in the 1960s and 1970s. Leiden: Brill, 1982.
- Barnes, Samuel H.; Kasse, Max; Allerback, Klaus R.; Farah, Barbara G.; Heunks, Felix; Inglehart, Ronald; Jennings, M. Kent; Klingemann; Hans D., Alan; and Rosenmayer, Leopold; Political Action: Mass Participation in Five Western Democracies. Beverly Hills: Sage Publications, 1979.
- Dogan, M. and Pelassy, D., How to Compare Nations. Chatham, N.J.: Chatham House, 1984.
- Hippler, Hans-J. and Schwarz, Norbert, "Not Forbidding Isn't Allowing: The Cognitive Basis of the Forbidding-Asymmetry," Public Opinion Quarterly, 50 (Spring, 1986), 87-96. Allow
- Inglehart, Ronald and Rabier, Jacques-Rene, "If You're Unhappy, This Must be Belgium: Well-Being Around the World," Public Opinion, 8 (April/May, 1985), 10-15.
- Inglehart, Ronald, The Silent Revolution: Changing Values and Political Styles Among Western Publics. Princeton: Princeton University Press, 1972.
- Jackman, Robert W., "Cross-National Statistical Research and the Study of Comparative Politics," American Journal of Political Science, 29 (Feb., 1985), 161-182.
- Jasper, James M., "Two or Twenty Countries: Contrasting Styles of Comparative Research," Comparative Social Research, 10 (1987), 205-229.
- Kohn, Melvin L., "Cross-National Research as an Analytic Strategy," American Sociological Review, 52 (December, 1987), 713-731.
- Kuechler, Manfred, "The Utility of Surveys for Cross-National Research," Social Science Research, 16

(Sept., 1987), 229-244.

Niessen, M. and Peschar, J. (eds.), International Comparative Research: Problems of Theory, Methodology, and Organization in Eastern and Western Europe. Oxford: Pergamon, 1982.

Przeworski, Adam and Teune, Henry, The Logic of Comparative Social Inquiry. New York: Wiley-Interscience, 1971.

Rokkan, Stein; Viet, Jean; Verba, Sidney; and Almsay, Elina, Comparative Survey Analysis. The Hague: Mouton, 1969.

Smith, Tom W., "The International Social Survey Program," Journal of Official Statistics, 2 (1986), 337-338.

Smith, Tom W., "The Polls: The Welfare State in Cross-National Perspective," Public Opinion Quarterly, 51 (Fall, 1987), 404-421.

Smith, Tom W., "Social Inequality in Cross-National Perspective," GSS Cross-National Report No. 7. NORC: Chicago, Oct., 1988.

Szalai, Alexander and Petrella, Riccardo, eds., Cross-National Survey Research: Theory and Practice. London: Pergamon Press, 1977.

Treiman, Donald J., Occupational Prestige in Comparative Perspective. New York: Academic Press, 1977.

Vallier, Ivan, ed., Comparative Methods in Sociology. Berkeley: University of California Press, 1971.

Verba, Sidney; Nie, Norman; and Kim, Jae-On, The Modes of Democratic Participation: A Cross-National Comparison. Beverly Hills: Sage Publications, 1971.

Verba, Sidney; Nie, Norman; and Kim, Jae-On, Participation and Political Equality: A Seven-Nation Comparison. Cambridge: Cambridge University Press, 1978.