RESEARCH · RESEARCH DESIGN

Comparative Research Across Countries and Cultures

How comparative research across countries and cultures works, the logic of most similar and most different designs, the problem of equivalence in concepts, translation, samples and measurement, cultural response styles, and practical steps for producing comparisons that are genuinely comparable.

ModQu research article · 1,447 words · 15 references (APA 7th edition)

Comparison is fundamental to research. Knowing that 40% of customers in one country prefer to shop online means little until it is compared with other countries, other periods or other groups. **Comparative research** across countries and cultures makes such comparisons systematically, asking how and why societies, organisations and people differ. It underpins major research programmes on values, leadership and institutions, and it is essential for any organisation that operates in more than one market. It is also unusually difficult. A question that means one thing in Toronto may mean something different in Tokyo, a sample drawn in one country may not be comparable with one drawn in another, and people in different cultures may use rating scales in different ways. This article explains the logic of comparative research, the main threats to comparability and the practices that address them.

Why compare?

Comparison serves several purposes. It shows whether a finding is general or specific to one setting. Henrich et al. (2010) documented that most behavioural research has studied people in Western, educated, industrialised, rich and democratic societies, and showed that findings once assumed to be universal, from visual perception to fairness in economic games, vary substantially across populations. Comparison also allows researchers to test explanations. If a characteristic of a society is thought to cause an outcome, societies that differ in that characteristic should differ in the outcome. And comparison provides benchmarks: an organisation's performance in one market is easier to interpret alongside its performance elsewhere.

The logic of comparison

Comparative research often deals with a small number of cases, such as a handful of countries, that differ in many ways at once. Mill (1843) set out two methods that remain the basis of comparative design. The **method of agreement** compares cases that share an outcome but differ in most other respects, looking for the one factor they have in common. The **method of difference** compares cases that are similar in most respects but differ in the outcome, looking for the factor that distinguishes them.

Przeworski and Teune (1970) translated these into two research strategies. A **most similar systems design** selects countries that are alike in many respects, such as Canada and Australia, so that the differences between them can be more readily linked to the few factors on which they differ. A **most different systems design** selects countries that differ widely, so that any relationship that holds across all of them is likely to be general. Lijphart (1971) noted the central difficulty of comparative research: many variables and few cases. Careful case selection, and focusing on the variables that matter most, are the main responses.

Ragin (1987) argued that comparative researchers should treat each country as a configuration of conditions rather than as a collection of independent variables, because the effect of one factor often depends on the presence of others. A policy that works in a country with strong public institutions may fail in one without them.

Large-scale comparative studies

Some of the best-known research in management and social science is comparative. Hofstede (2001) analysed surveys of IBM employees in more than 50 countries and identified dimensions on which national cultures differ, such as individualism and the acceptance of unequal power. Because all respondents worked for the same company in broadly similar jobs, the design held many factors constant, a version of the most similar systems logic. The GLOBE study (House et al., 2004) surveyed managers in 62 societies about cultural practices, values and effective leadership. Schwartz (1992) developed a theory of basic human values and tested its structure with samples from 20 countries, finding a similar structure of values across very different cultures, alongside substantial differences in the priority given to each value.

These studies have been influential and also widely debated, largely because of the methodological challenges discussed below.

The problem of equivalence

The central methodological issue in comparative research is **equivalence**: whether what is being measured is the same thing, measured in the same way, in each setting. van de Vijver and Leung (1997) distinguish several levels.

Is the comparison comparable?

Four kinds of equivalence needed before countries can be compared

Equivalence	The question it asks	Common remedy
Conceptual	Does the concept mean the same thing in each culture?	Qualitative groundwork in each country
Translation	Do the questions convey the same meaning in each language?	Team translation, back-translation, pretests
Sample	Do the samples represent comparable populations?	Matched sampling methods; report coverage
Measurement	Do people use the scale and items in the same way?	Invariance tests; adjust for response styles

Based on van de Vijver & Leung (1997); Harkness et al. (2003) ModQu.com

Figure 1. Four kinds of equivalence that a cross-national comparison needs, with the question each asks and a common remedy. Based on van de Vijver and Leung (1997) and Harkness et al. (2003); table by ModQu.

Conceptual equivalence

A concept must have a comparable meaning in each culture before it can be compared. "Loyalty" to a brand, "trust" in institutions or "satisfaction" with a job may be understood differently, or may not be a meaningful category everywhere. Preliminary qualitative research in each setting helps establish whether a concept travels.

Translation equivalence

Questions must convey the same meaning in each language. Literal translation often fails, because words carry different connotations. Brislin (1970) proposed **back-translation**: one translator converts the questionnaire into the target language, a second translates it back without seeing the original, and differences are examined and resolved. Harkness et al. (2003) recommend going further, with a team approach in which several translators, a reviewer and a subject expert discuss and agree each item, followed by pretesting with respondents in each country.

Sample equivalence

Samples must represent comparable populations. Comparing a national probability sample in one country with an online panel of urban internet users in another will mix national differences with differences in sampling. Even good samples may differ in coverage, response rates and the way respondents were contacted.

Measurement equivalence

Even when questions are well translated, respondents in different countries may interpret or use the response scale differently. Steenkamp and Baumgartner (1998) set out statistical tests of **measurement invariance**, which check whether the relationships between items and the underlying construct are the same across groups. Vandenberg and Lance

(2000) describe a sequence of increasingly strict tests. Only if the strictest level, **scalar invariance**, holds can average scores be compared directly across countries. Without it, a difference in average satisfaction might reflect a difference in how people use the scale rather than a difference in satisfaction.

Response styles

One reason measurement equivalence fails is that cultures differ in **response styles**: systematic tendencies in how people answer questions regardless of content. Baumgartner and Steenkamp (2001) studied consumers in eleven European Union countries and found substantial national differences in acquiescence (the tendency to agree), extreme responding (choosing the ends of scales) and midpoint responding. Johnson et al. (2005) examined data from 19 countries and found that extreme responding was more common in cultures that value individualism and masculinity, while acquiescence was more common in cultures that value collectivism and avoid uncertainty.

These differences can produce misleading comparisons. If respondents in one country tend to choose the highest point on any scale, their satisfaction scores will look higher than those of a country where respondents avoid extremes, even if underlying satisfaction is the same.

Practical steps

Several practices improve the comparability of cross-national research.

- **Start with the concept.** Use qualitative research in each setting to check that the concept is meaningful and to identify how it is expressed locally.
- **Translate as a team.** Combine forward translation, review, back-translation where useful, and pretesting with real respondents in each language (Harkness et al., 2003).
- **Match samples.** Use comparable sampling methods and populations, and report differences in coverage and response.
- **Test for invariance.** Before comparing average scores, check that the measure works the same way in each country (Steenkamp & Baumgartner, 1998).
- **Measure and adjust for response styles.** Include items that allow response styles to be estimated, use rankings or choice tasks where appropriate, and compare patterns within countries as well as levels across them.
- **Use local expertise.** Involve researchers from each country in design and interpretation, so that results are read in context.

Implications for international market research

Organisations comparing customer satisfaction, brand health or employee engagement across countries face all of these issues. A global satisfaction league table that ranks countries on average scores may say more about national response styles than about the customer experience. Comparing changes over time within each country, or comparing the ranking of attributes within each country, is often more informative than comparing raw scores across countries. Where cross-country comparisons of levels are essential, they should be supported by evidence that the measure is equivalent, and reported with appropriate caution.

Conclusion

Comparative research across countries and cultures shows whether findings are general or local, tests explanations that depend on context, and provides benchmarks for international decisions. Its value depends on equivalence: comparable concepts, translations, samples and measurement. Careful design, team translation, invariance testing and attention to response styles can make comparisons meaningful. Without them, apparent differences between countries may reflect nothing more than differences in method.

References

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(2), 143–156. https://doi.org/10.1509/jmkr.38.2.143.18840

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185–216. https://doi.org/10.1177/135910457000100301

Harkness, J. A., van de Vijver, F. J. R., & Mohler, P. Ph. (Eds.). (2003). *Cross-cultural survey methods*. Wiley.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). SAGE.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. SAGE.

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36*(2), 264–277. https://doi.org/10.1177/0022022104272905

Lijphart, A. (1971). Comparative politics and the comparative method. *American Political Science Review, 65*(3), 682–693. https://doi.org/10.2307/1955513

Mill, J. S. (1843). *A system of logic, ratiocinative and inductive*. John W. Parker.

Przeworski, A., & Teune, H. (1970). *The logic of comparative social inquiry*. Wiley-Interscience.

Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. University of California Press.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology, 25*, 1–65. https://doi.org/10.1016/S0065-2601(08)60281-6

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78–90. https://doi.org/10.1086/209528

van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. SAGE.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002