Cross-Country Comparisons of Cultural Statistics: Issues and Good Practice

Christopher Madden 2005

Abstract

Cross-country comparisons are popular in cultural policy. This paper looks at how cultural statistics are used in the making of such comparisons. Analysts have identified a general 'sloppiness' in current cultural statistics comparisons. Some of the major problems in both data production and data presentation are documented, and a 'checklist' of good practice is provided. The paper aims to provide guidance and ideas for anyone making cross-country comparisons of cultural statistics.

Keywords: Cultural statistics; Cultural indicators; International comparisons; Comparative cultural policy

Introduction

Comparative policy analysis—the comparing of policies between and across countries—is a popular analytical tool in cultural policy. Cultural policy analysts, researchers and commentators regularly compare cultural sectors and cultural policies between countries, draw policy inferences from one country to another, and make prescriptions and recommendations based on the experiences of ‘other’ countries. Arguments are often based on comparative cultural statistics or indicators.

This paper focuses on the use of cultural statistics in such comparative analyses. After brief reviews of the current state of cultural statistics and comparative cultural policy analysis, the paper reviews the cultural policy literature to investigate three basic questions: How well are cross-country comparisons made? What are the pitfalls and problems in making comparisons? And what are some of the solutions to these problems?

1 This is an electronic version of an article published in Cultural Trends, Volume 14, Issue 4 December 2005, pp.299-316. Available online at:
The aim of the paper is to provide guidance and ideas to anyone undertaking cross-country comparisons of cultural statistics. The main findings are summarized in the Appendix as a ‘checklist’ of good-practice techniques taken from the cultural policy literature.

The paper concentrates almost exclusively on the cultural policy literature. There are two other research areas of particular relevance to the topic: the literature on statistical comparisons in other policy domains (such as health, education and economics); and the literature on policy transfer methodologies (see in particular the references and discussion in Dewey & Wyszomirski, 2004, but also Nedley, 2000, and Wyatt & Grimmeisen, 2002). For reasons of manageability, this literature was not consulted for this paper.

This work was commissioned by the Australia Council for the Arts as part of the Council’s ongoing research programme (Australian Council, 2004).

**Part 1: Cultural Statistics**

While there has been a surge in the availability of cultural statistics since the 1980s, many cultural policy researchers and policy makers are questioning both the quality and the policy relevance of the data (see e.g. Allin, 2000; European Taskforce on Culture and Development, 1997; Schuster, 2002a, 2002b; Selwood, 2002; Wiesand, 2002).

In his analysis of the ‘state of play’ in cultural statistics, Schuster (2002a) argues that:

- there has been a ‘dramatic’ resurgence of investment in information and research in the field of cultural policy;
- much of the cultural policy research that is being undertaken around the world today might better be thought of as the development of a statistical base of data rather than the conduct of policy-relevant research;
- it is not so much the shortage of data that should command one’s attention—rather, it is the lack of use of those data that needs to be addressed;
- with the proliferation of data from a wide variety of sources, the issue of how to assure the quality of the data has become even more important.

Selwood echoes many of Schuster’s observations:

> [A]s a result of initiatives going back to the 1980s we have by now accumulated a mass of primary data. But how valuable has the pursuit of that data for the cultural sector actually been? Whatever the discussions about the quality of evidence gathered, questions also need to be asked about the use, or the lack of use, to which it is put ... Until the cultural bureaucracy’s analysis of data is guaranteed, and until the evidence gathered can be seen to be being used constructively, it could be argued that collecting data has been a relatively spurious exercise. (2002, p. 12)

\[^2\text{The increase in the production of cultural data appears to be confined to developed nations. Goldstone (1998, p. 350) suggests that the lack of basic cultural statistics in the ‘poorest countries’ is ‘crippling’ (see also Development Gateway, 2002; Glade, 2003).}\]
To redress the perceived imbalance between the supply and the use of cultural data, much work is now being undertaken to focus the production of cultural statistics more strategically—rather than simply creating data for their own sake, analysts are reviewing what it is that cultural statistics are needed for and how best to target data production to those needs. Statisticians are also assessing in greater detail what it is that cultural variables are measuring, and whether these data make sense in cultural terms (Foote, 2002; Guetzkow, 2002; Matarasso, 2001; McCarthy & Jinnet, 2001; McCarthy, Ondaatje, & Zakaras, 2001; Princeton University, 2002; Selwood, 2002; Yoshitomi, 2002). Some analysts doubt that statistics can ever measure all of the policy-salient aspects of culture, especially if data are to be used for making cross-country comparisons (McCaughey, 2002; Price, 1999; Sen, 1998).

This process of review is driven by two notable international agendas. The first is ongoing work to develop meaningful ‘cultural indicators’ (International Federation of Arts Councils and Culture Agencies, IFACCA, 2005). The second is the aspiration to promote international standardization in cultural statistics to feed an ever-growing demand for comparative cultural policy analysis (European Commission, 2000; Lievesley, 2001; Manninen, 2002; Walker, Scott-Melnyk, & Sherwood, 2002).

A lack of good quality, policy-relevant cultural data appears to have prompted a flurry of reviews about what cultural statistics should be measuring and how such statistics should be generated. This is most evident in work being undertaken to review and reshape cultural statistics ‘architectures’—the statistical frameworks, definitions and classifications used to describe art and culture. The variety of topics and range of countries represented at the International Symposium on Culture Statistics in Montréal in October 2002 is testament to the substantial development work currently being undertaken in cultural statistics.³

Such development work will undoubtedly lead to more readily comparable cultural statistics in the future. At present, however, the state of flux caused by development work hampers inter-country comparisons, as innovations in cultural statistics do not occur uniformly across countries.

In summary, there is currently a steady supply of cultural data in developed nations, although concerns over the quality of these data have prompted substantial work to improve data by developing better methodologies and statistical architectures. Two crucial implications for the current project are:

1. There is a concern over the quality of cultural data—cultural statistics should be viewed with a healthy scepticism;
2. Standardization is still a work in progress—there are substantial limitations to comparing cultural data between countries.

Part 2: Comparative Cultural Policy Analysis

Comparative cultural policy analysis is an important element of much cultural policy analysis and research, and has been the subject of lively academic debate (Belfiore,

---

³ Symposium papers are available on-line at www.uis.unesco.org/ev.php?ID=5509_201&ID2=DO_PRINTPAGE
A number of reasons have been put forward for cultural analysts’ desire for international comparison:

- An innate suspicion of national, and potentially ‘nationalistic’, cultural policies, meaning that policy development can only be fully trusted in the context of independent verification (Schuster & Amad, 2002);
- Increasing global integration of cultural practices, which makes internationally focused analysis superior to nationally focused analysis (Belfiore, 2004; Foote, 2002; Schuster & Amad, 2002);
- Increased global integration of nation states and the rise of cross-national governance (Jowell, 1998; Schuster, 2002b, p. 35);
- A vagueness in cultural policy, brought about by the esoteric nature of culture, which promotes comparison as a way of contextualizing and articulating abstract cultural policy issues (Wiesand, 2002);
- A dearth of quality data, discussed earlier, which causes policy analysts to seek overseas information for adapting to local conditions, or for double checking untrusted local data (Wiesand, 2002).

Lievesley’s (2001, p. 378) reasons for developing comparative cultural statistics mirror the policy imperatives above:

- To enable countries to gain a greater understanding of their own situation by comparing themselves with others, thus learning from one another and sharing good practice;
- To permit the aggregation of data across countries to provide a global picture, thus enabling the design of international initiatives informed by evidence;
- To provide information for nation-state accountability, and for the assessment, development and monitoring of supranational policies.

Jowell simply states that ‘[t]he importance and utility to social science of rigorous cross-national measures is incontestable’ (1998, p. 168).

The surveys in Kelland and Selwood (2002) and Schuster and Amad (2002) indicate that cultural statisticians ply a healthy trade in international comparisons. Schuster is optimistic about such comparisons:

Methodology has become sufficiently refined over time and sufficiently similar across countries that, for the first time, it has become possible to envision truly cross-national comparative studies. (2002b, p. 21)

Belfiore takes a more pessimistic view:

The methodologies that currently guide comparative cultural policy research are largely inappropriate, and do not meet the specific requirements of cross-national research. (2004, p. 41)

The next section explores some of reasons why caution should be exercised when making cross-country comparisons of cultural statistics.
Part 3: Cross-Country Comparisons: Limitations and Problems

A rise in the production of cultural statistics in a number of countries has made the prospect of international statistical comparison a real possibility, and a substantial number of comparative studies already exist. Yet the cultural policy literature is peppered with warnings about international data comparisons (see e.g. Belfiore, 2004; European Taskforce on Culture and Development, 1997; Glade, 2003; IFACCA, 2002; Lievesley, 2001, p. 378; Mitchell, 1996a, 1996b; Price, 1999; Sen, 1998).

With apparent hypocrisy, stern warnings about comparisons are sometimes made by analysts who are themselves undertaking a comparison.

There is a long list of documented problems in making inter-country comparisons of cultural data. A good introduction can be found in Mitchell (1996a, 1996b). For simplicity, the problems highlighted in this working paper will be grouped into two types: data production problems and data presentation and utilization problems.

Data Production Problems
Data supply issues present the most obvious barriers to making cross-country data comparisons. Some of the more common supply issues are listed below.

Data availability
As already noted, although many developed countries have amassed a substantial base of cultural data, data availability is still lacking in many countries. This limits the scope of comparative analyses.

Data quality
As noted earlier, concerns have been expressed over the quality of cultural data. Issues of quality identified include: inherent difficulties in quantifying culture; methodological shortcomings in extant data; and unreliability of data generated in self-serving research by cultural agencies (Goldstone, 1998; Nielson, 1999; Schuster, 2002b, p. 21; Selwood, 2002; van Puffelen, 1996). If cultural data reflect poorly the cultural reality within countries, problems are exacerbated when comparing data between countries.

Non-standardization of data
Despite a rising standardization in cultural statistics, it is still extremely rare for cultural data to be gathered in exactly the same way in different countries. It is useful to consider two types of differences: differences in definitions and classification, and differences in methodologies.

Differences in definitions and classifications
Different countries have different definitions of culture (i.e. differences in the scope or boundaries of culture) and different classifications of culture (i.e. the cultural domain is broken down in different ways). Many of these fundamental differences result from differences in institutional arrangements in the cultural policy arena (for example, some cultural ministries have responsibility for the design sector, others do not, and their definitions will reflect this difference). Other differences may be philosophical, theoretical, historical or simply arbitrary. However they arise, differences in
definitions reduce the degree of comparability of cultural data. In the case of arts participation surveys, for example, Pronovost notes that ‘there is little consensus on the international level as to the choice of categories of cultural activities,’ so that ‘identifying an activity as simple as reading is a problem’ (2002, p. 3). Schuster’s (1987) three approaches to dealing with issues of boundary definition in comparative analysis—inclusive, floating and anchored boundaries—are now part of the comparative cultural policy lexicon. The anchored boundary approach, where the same definition is applied across different countries, is clearly the most attractive from a statistical point of view as it ensures that ‘like is compared with like’. However, it is not always possible to apply the same definition to data from different countries, especially where data are from published sources. It is not surprising, then, that the floating boundaries approach, where the definition of the arts or culture is allowed to vary between countries, is the most common in cross-country comparisons of cultural data. But this approach places additional importance on clear, informative presentation (as discussed later).

Differences in methodologies
There are few surveys that use the same survey questionnaire and format across countries (a recent exception is the Eurobarometer survey of Europeans’ participation in cultural activities; Spadaro, 2002). Countries more commonly design their own survey instruments, which means that methodologies also tend to differ. Although surveys may measure broadly similar cultural phenomena, such as arts participation, more often than not the surveys are for different years, different time frames (e.g. last four weeks, previous 12 months) and different age groups, and use different question formats, wording of questions and survey instruments (e.g. phone, face-to-face or postal surveys; census or sample populations). Even with commonplace variables that might appear highly comparable, such as cultural participation rates, methodological differences can make comparison difficult or impossible. The slightest difference in methodology can prevent a ‘true’ comparison.

Attempts to account for methodological differences are fraught with difficulty. An adjustment may require the recalculation of a survey’s raw data, which is not always practical; or it may require the estimation of a counterfactual, which is prone to inaccuracy. A valid adjustment may simply not be possible.

Structural inter-country differences
Even where data are produced in similar ways, structural differences between nations can cause biases that complicate inter-country comparisons.

An obvious example is the impact of institutional and policy differences on the efficacy of comparisons of government arts expenditures. Governments in different countries can have immeasurably different approaches to cultural policy. Some countries have an arm’s length arts council, others a ministry for the arts or culture. In some countries arts education policies and programmes are the domain of a ministry of culture; in others they are the responsibility of a ministry of education. Different countries have different degrees of decentralization in cultural provision and different levels of government administration. Few countries have the same rules governing charitable donations, tax incentives and artists’ earnings. Schuster (2002a, p. 6) points out more structural differences that can influence data on government arts expenditure.
Cultural differences between countries can also bias data, even when data are generated from the same survey. For example, in the Eurobarometer survey of Europeans’ cultural participation, problems were experienced in comparing national data due to differences in local interpretations of apparently familiar terms such as ‘concert’, ‘opera’ and ‘museum’ (Skaliotis, 2002).

Structural differences such as these place significant additional methodological pressures on cross-country comparisons: first, on the pre-selection process of determining which countries to compare (Ebbinghuas, 2005); and second, once countries have been chosen, on methodological contextualization, or accounting for the different ‘social, cultural and political contexts in which social phenomena manifest themselves’ (Belfiore, 2004, p. 44).

At first sight it might appear that many of the problems outlined above could be avoided through better standardization of cultural data. However, it appears that the problems are present even in cross-national surveys—i.e. single surveys that are applied across a number of countries. Jowell records a long list of problems in cross-national studies, using as a test case the International Social Survey Programme, which is ‘widely regarded as a conscientious, rigorous, and successful model of a cross-national social survey’, and for which, ‘despite its high ambitions and the relative diligence with which it applies its standards and rules, inconsistencies between countries still abound’ (Jowell, 1998, p. 196). Similar sentiments can be found in Skaliotis (2002) and Spadaro (2002) for the Eurobarometer survey on Europeans’ participation in cultural activities. Lievesley pessimistically suggests that ‘[n]o cross-national study can be perfect as far as comparability is concerned and it is only too easy to find reasons why the data should not be taken seriously’ (2001, p. 381).

The data production problems discussed here are hardly unique to culture—they are the curse of comparative statisticians in all areas of public policy. The issue for comparative cultural research is whether these problems are significant enough to make comparisons futile or ill-advised. Glade takes a pessimistic view:

So varied, from country to country, are the definitions of the cultural sector and its components and so varied the methodologies applied to its study, starting with data collection and classification, the task of finding data with sufficient comparability to be useful in comparative analysis is monumental. (2003, p. 23)

---

4 Here cultural differences between countries are being blamed for causing problems in comparing data that measure culture. This paradox, endemic to much cultural analysis, is unravelled in detail in McHoul (2004).

5 Similar problems are noted by researchers in policy areas with more established statistical traditions than culture. Examples are for: crime statistics (Barclay, Tavares, & Siddique, 2002); labour force statistics (Capdevielle & Sherwood, 2002); and agricultural industry statistics (American Association of Agricultural Economics, AAEA, 2000).
Data Presentation and Utilization Problems
There are indications of a general sloppiness in the presentation of comparative cultural data (IFACCA, 2002; Kelland & Selwood, 2002). Some of the major issues are set out below.

Presentation without context
The most common concern is that data are presented without proper context. Lievesley finds that ‘[h]ow indicators are presented can be a cause for concern. They are often stripped of their essential metadata’ (2001, p. 378). As UNESCO notes: ‘[p]roper metadata systems provide information on methodology required in order to ascertain whether the data set in question is fit for use in cross-national comparisons’ (2003, p. 51). As Belfiore (2004) points out, contextualization can extend beyond simply describing statistical metadata to include accounts of differences and similarities in the social, cultural and political systems of countries being compared. A further problem is the lack of understanding and accounting for ‘exogenous’ influences on the statistical indicators used in cross-country comparisons. IFACCA (2005), for example, discusses the impact on performing arts attendance rates of exogenous factors such as relative price differentials, disposable incomes, demographic shifts and programming differences. Accounting for these exogenous factors will provide a clearer indication of the actual differences in arts involvement between populations in different countries; however, few, if any, comparisons of cultural data aspire to such a level of sophistication.

Presentation with obscured context
More sophisticated analyses attempt to account or adjust for underlying differences in data, or if differences cannot be reconciled, at least note these irreconcilable differences as caveats (some examples of these more detailed analyses are: European Commission, 2000; Feist, 1998; International Arts Bureau, 2000). However, IFACCA finds for many of these more carefully constructed analyses that ‘caveats are typically buried within the text, while data remain boldly tabulated side-by-side. Differences underlying the data are thus obscured, and too often conclusions are drawn from incomparable data’ (2002, p. 6).

Although many cross-country tabulations do not pretend that data are comparable, the very act of presenting data in the same table only encourages comparison and therefore misuse.

Lack of sensitivity analysis
Given the long list of data production differences noted in the previous section—including different definitions, classifications and statistical methodologies—it would be advisable for those making comparisons to undertake sensitivity analyses. Sensitivity analyses would try to quantify the impact of such differences on the cross-country data presented. A good example of a sensitivity analysis occurs in International Arts Bureau (2000), which calculates how sensitive measures of government arts expenditure are if indirect expenditure is accounted for as well as direct expenditure. The analysis estimates that tax foregone through the Irish Government’s arts tax instruments increases government arts expenditure by 50

6 ‘Metadata’ include such information as data dictionaries, record layouts, questionnaires, sample designs and standard errors (United Nations Educational, Scientific and Cultural Organization, UNESCO, 2003, p. 52).
percent (from 12.36 to 19.22 Irish Pounds per capita). Similar estimates are not able to be calculated for other countries in the comparison, and herein lies the main problem of cross-country sensitivity analyses: it is often very difficult to obtain enough information to be able to perform a sensitivity analysis for each country in a comparison, especially where data are from secondary sources.

League tables
League tables are so common and so controversial that they deserve a special note. The league table is the archetypal cross-country comparison: it is a data table that ranks countries by some chosen variable (per capita government arts expenditure is a particularly popular league table in cultural policy). Schuster notes:

Through the 1970s and the 1980s, the league table became the sine qua non [i.e. the essential element] of much comparative research on arts funding. In the mid 1980s I counted sixteen different studies in English that had generated sixteen different such tables and was able to demonstrate that, depending on the methodologies and the biases of each of the studies, countries jumped all over the time series graphs on which I compared the results of these studies. Their popularity continues unabated, though subsequent research has made it more and more difficult for researchers to remain ignorant of the issues [raised earlier in Schuster’s paper]. Nonetheless, the literature is full of such tables; they are very hard to resist. (2002a, p. 7)

Schuster demonstrates the dangers of the league table using the example of arts funding per capita. He concludes that the flaws of league tables are simply ‘apparent’. Lievesley agrees, expressing grave concerns over the way in which league tables are constructed, presented, manipulated and misused. She concludes that ‘it is debateable whether the creation of these league tables is a statistical exercise’ (Lievesley, 2001, p. 378).

Misuse, misinterpretation and other ‘strategic hazards’
A common concern about international comparisons of data—and league tables in particular—is in how the comparisons are interpreted and re-used by others. Cautious researchers may provide a full list of warnings and caveats about data limitations, but they have little control over how others use their work. Even the most carefully tabulated, fully caveated, comparative analyses can be misinterpreted, used out of context, reproduced without caveats, manipulated or otherwise abused. Schuster illustrates one such ‘strategic hazard’:

In 1984–85 I was asked by the Policy Division of the National Endowment for the Arts to conduct a comparative study of the structure and level of funding for the arts and culture in eight countries in Western Europe and North America ... In the latter stages of my research I received a call from a Canadian government researcher. His minister had passed along an emergency request. He needed to know how Canada stacked up against others in arts funding, and he needed to know yesterday. This researcher pleaded with me to release my preliminary figures to him. His job would be so much easier if he

---

7 Such misuse can be thought of as a ‘strategic hazard’, a generalized form of the ‘moral hazard’ well-documented in law and economics (see http://en.wikipedia.org/wiki/Moral_hazard) (retrieved June 22, 2005).
could take advantage of work that had already been done. I relented, and we spent considerable time going over the numbers and my lengthy list of methodological caveats and footnotes. Within a day or two said minister was on national television citing new comparative research showing that per capita expenditures for culture in Canada were at a high level, on a par with Sweden, France, Germany, and the Netherlands. I no longer have perfect recall of these ensuing events, but I seem to remember that the Canadian figure had managed to increase, rather substantially, overnight. (2002a, p. 6)

Lievesley describes other hazards of which the comparative statistician should be wary:

[F]requently politicians or senior civil servants are dismayed by the relative results for their countries and tend to blame the messenger rather than examining the message ... An outcome of this unhappy state of affairs is the withdrawal from cross-national research of those countries which achieve disappointing results, and occasionally a more catastrophic effect on the career of the national statisticians involved. (2001, p. 381)

Researchers and statisticians tend to covet clarity and objectivity, and are prone to disappointment when these principles are overlooked by others who may not share their standards of excellence. Although they may be unable to control others' actions, they would be wise to anticipate the vested interests for which their findings may be appropriated and try to take precautions to prevent the misuse. Equally, though, statisticians can beneficially exploit the strategic environment in which their research is communicated and interpreted. For example, Lievesley notes that, for development indicators, ‘international statisticians have sometimes been reluctant to elaborate on the limitations of league tables because they can see that the media and political attention given to league tables raises the profile of development’ (2001, p. 378).

Strategic issues such as these reinforce the importance of being clear about the objectives of comparison: the desired ends are often in conflict with the statistical means. Before comparison is made, it should be clear to what end the comparison is to be put, and this should be clearly communicated as an integral part of the research.

Part 4: Mitigating the Problems of Comparison

Jowell (1998, pp. 174–176) outlines 10 ‘rules of thumb’ for mitigating some of the problems of inter-country data comparisons, including the following.\(^\text{5}\)

- Compare familiar countries: ‘Social scientists should undertake not to interpret survey data relating to a country about which they know little or nothing’.
- Limit the number of countries: ‘Analysts of cross-national data should resist the temptation to compare too many countries at once’.

\(^{5}\) It should be remembered that Jowell is focusing on cross-national surveys, i.e. on single surveys applied across a number of countries. However, the principles apply equally well to comparisons of independent national surveys.
State limitations clearly: ‘Social scientists contemplating or engaged in cross-national studies should be as open about their [study’s] limitations as they are enthusiastic about their [study’s] explanatory powers’.

Apply the strictest of quality controls: ‘[T]he same technical standards as we would impose for national surveys should apply to cross-national surveys too. Arguably, they should be even higher in view of the fact that country will inevitably be one of the major independent variables in the analysis.’

Be sceptical: ‘Analysts of cross-country data should undertake to suspend belief initially in any major intercountry differences they discover’.

Be aware of data context: ‘All too often, analysts seem to compare national data sets in vacuo [i.e. in isolation]’. This is more than simply understanding differences in frameworks and methodologies, it is also about considering the broader context within which a comparison is being made. Take, for example, comparisons of government expenditure on the arts: these comparisons cannot account for the many indirect forms of government support, such as tax relief for artists, tax incentives for arts consumers (e.g. consumption tax exemptions), tax incentives for philanthropic giving, and the variety of other forms of indirect support that governments provide to the arts. It is misleading to use direct government arts expenditures as an indicator of overall levels of government support for the arts without accounting for indirect forms of support.

Lievesley expresses doubt about how achievable the first two are in practice, but declares the last three ‘very important lessons for international statisticians’ (2001, p. 380).

A number of other recommendations might be added to Jowell’s list in the light of the discussion here.

- Have clear objectives: be clear about why a comparison is being made.
- Be clear about how to interpret data: if ‘real’ inter-country differences in data are found (rather than differences brought about by measurement biases), have a clear idea about how to interpret the differences and what the differences imply for policy. Be sure that data differences are not the result of ‘exogenous’ influences (for more on this see IFACCA, 2005, p. 17).
- Avoid ‘league tables’.
- Use trend data: trends will be less influenced by methodological and contextual differences, as long as these differences are stable over time. For example, it may be unwise to compare levels of government support between countries for any particular year, but trend data may provide a more reliable indication of which countries’ support is increasing and which countries’ support is decreasing (assuming that structural changes and other such exogenous influences can be accounted for).
- Use ratios and statistical indicators: ratios and other statistical measures can circumvent many of the problems associated with comparisons. For example, rather than compare artists’ incomes between countries, the ratio of artists’ incomes to average incomes might be used as a comparison statistic. The ratio allows implications to be drawn out about artists’ relative earnings between countries while avoiding the need to choose an appropriate real exchange rate.
Choice of exchange rate is also an issue when government per capita arts expenditure is being compared.

Schuster (1987, p. 2) proposes a set of 13 key questions that should be ‘kept in mind’ when undertaking, evaluating or reading international comparisons. These 13 questions are as relevant today as they were in 1987.

1) What appears to be the central research question in the study?
2) What is really at issue? Is there an underlying political agenda?
3) Is the research intended to compare or to explain? Do the methodological choices made by the researcher facilitate or frustrate that goal?
4) What is the boundary of analysis? Is it clear?
5) Is the boundary of analysis appropriate to the research?
6) How was the choice of countries made?
7) Are [the countries] appropriate to the research?
8) How were the data collected?
9) What were the data sources? Are they reliable?
10) Are the financial calculations made in agreement with the boundaries of analysis?
11) What summary statistics are used? Are they appropriate to the research?
12) Are the conclusions justified by the evidence?
13) Are the conclusions modestly presented, recognizing the limitations of the research method, the data and the analysis?

These rules of thumb and analytical questions serve as a good evaluative tool before and after undertaking cross-country comparisons. They are summarized in the Appendix as a rough ‘checklist’.

**Part 5: Summary**

Despite significant advances in the production of cultural statistics, our ability to make inter-country comparisons of cultural data is still limited. Data are generated in different ways and for different reasons. Differences in social and institutional environments between countries are difficult or impossible to account for. Structural differences between countries can significantly bias data. Many of these problems are present even in cross-country data that are generated from the same survey.

Cross-country differences in cultural data may, therefore, be due to differences in measurement rather than to ‘real’ differences in cultural phenomena. Benchmarking, or drawing policy and programme implications from such data, should be strongly discouraged.

On the other hand, there are a number of real benefits from undertaking comparative cultural policy analysis. Comparative statistical analyses can uncover information that is indispensable for domestic policy formulation and evaluation. The list of problems documented here may seem overwhelming, but measures can be taken to mitigate their impact on comparative data and allow limited, cautious comparisons to be made. But, if the decision is made to undertake a cross-country comparison despite all the
drawbacks, pitfalls and strategic hazards, there is one crucial question that needs to be clearly thought out and articulated: why is a comparison being made at all?

Acknowledgements
This article is based on research commissioned by the Australia Council for the Arts (Australia Council, 2004). The author thanks Sarah Gardner, Emma Barron and Sarah Barns for guidance and suggestions. Opinions, errors and omissions are solely the author’s. The views expressed here are not necessarily those of the Australia Council, nor the board and members of the IFACCA.
References


Appendix. Checklist for Cross-Country Cultural Data Comparisons

Below is a selection of recommendations and issues to be considered by anyone wishing to undertake an international comparison of cultural data. These are taken from the literature and analysis in the previous pages, and grouped into three broad categories. Please feel free to suggest further items for the list by emailing info@ifacca.org.

(1) Planning the comparison
- Be sceptical
- Be clear about why is a comparison being made
- State the research question
- State the research objectives
- Clearly define the domain that is the basis for comparison (art, culture, cultural industries, creative industries, etc.) and its constituent elements (visual arts, music, etc.)
- Decide whether boundaries for the definition will be ‘inclusive’, ‘floating’ or ‘anchored’
- Consider how the comparison will be reported and used by others. Anticipate and consider how to avoid possible misuse

(2) Selecting countries and statistical variables
- Compare familiar countries
- Limit the number of countries being compared
- Use trend data, ratios and statistical indicators
- Use information sources as close as possible to the data source
- Assess possible statistical indicators to be used:
  - Are they unambiguous?
  - Are they appropriate for the research objectives?
  - What do changes in the indicator mean?
  - How sensitive are the data to differences in definitions, data architectures and statistical methodologies?
  - What external (or ‘exogenous’) factors might cause apparent cross-country differences in data?
- Consider other institutional and structural differences that might influence the comparison.

(3) Presenting the results
- Clearly state the reasons why countries are chosen
- Provide ‘metadata’ for each country’s collection
- Provide contextual information on relevant social, cultural and political aspects of the arts and culture for each country
- State data limitations clearly in a set of caveats
- Place caveats within—or as close as possible to—data tables and graphs
- State clearly how ratios and indicators are calculated
- Avoid league tables
- Avoid conjecture: only draw conclusions that are clearly supported by the data; do not draw conclusions that can be disputed from alternative interpretations of the same data.