

Ethnicity and the comparative analysis of contemporary survey data

Lambert, Paul S.

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Lambert, P. S. (2005). Ethnicity and the comparative analysis of contemporary survey data. In J. H. P. Hoffmeyer-Zlotnik, & J. Harkness (Eds.), *Methodological aspects in cross-national research* (pp. 259-277). Mannheim: GESIS-ZUMA. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-49162-2>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

ETHNICITY AND THE COMPARATIVE ANALYSIS OF CONTEMPORARY SURVEY DATA

PAUL S. LAMBERT*

1 Introduction

Concepts of minority ethnic group are important to a great many social science analyses. Ethnic differences are often a focus of analysis; even when they are not, a case can very often be made that interactions between ethnicity and the topic of study will be non-ignorable (Anthias, 2001). Yet developments in the sociological conceptualisation of ethnicity generate serious problems for cross-national survey researchers. Typically, sociological discussions have argued the need for ever finer qualitative differentiations between different ethnic situations in any particular country. The task of mapping and analysing the *minutiae* of such ethnic differences, in terms of comparable variables in cross-national survey datasets, becomes highly problematic.

This challenge is not, however, one which social researchers should shy away from. The internationally comparative analysis of ethnicity effects has considerable substantive appeal, as different outcomes might be related to variations in national contexts and policies (e.g. Castles, 1993). At a more practical level, history suggests that comparative analyses of ethnicity, using whatever survey data categories are available, will continue apace – regardless of any misgivings about the validity of the variable indicators available.

* The data analysed in this paper was gratefully provided by: European Social Survey; UK Data Archive; Economic and Social Data Service (UK); Luxembourg Income Study; International Public Use Microdata Service. Any errors in the subsequent use of these datasets are the author's responsibility. The research was assisted by an IRIS grant from the CEPS research institute.

2 Concepts of Ethnicity in Social Survey Research

Relatively few methodological writings reflect upon measures of ethnicity in cross-national surveys, and, when attention is paid, it is often restricted to an overwhelmingly critical perspective (Allen & Macey, 1990; Rea et al., 1999; Favell, 2003: 25-34). There are, however, a few prescriptive evaluations of the comparative analysis of ethnicity through micro-social surveys (e.g. Lambert & Penn, 2001; Hoffmeyer-Zlotnik, 2003; Bonifazi & Strozza, 2003). There are also several outputs based upon cross-national survey research (e.g. Stille, 1999; Heckmann et al., 2001; Evans & Need, 2002; SYSDM, 2003; Jacobs & Tillie, 2004; van Tubergen et al., 2004¹).

In a recent review of the measurement of race and ethnicity in comparative survey research, Hoffmeyer-Zlotnik (2003) highlighted the complications caused by lack of consistency in just which underlying concept different social scientists have in mind when they discuss race and ethnicity (see also Aspinall, 2002). This underlying concept of interest can be termed the '*referent*' of the ethnicity measure (cf. Lambert & Penn, 2001). In previous literatures, the referent chosen has ranged over topics such as citizenship, national origins, country of ancestral origins, racialised visibility, language spoken, subjective cultural identity, and religion. In this writing, we use '*ethnicity*' as an umbrella term to refer to differences associated with any of these concepts. This reflects the use of the term in British literature, as a self-assigned category free to incorporate diverse influences (cf. Hutchinson & Smith, 1996; Banton, 1997)². Table 1 describes the range of alternative ethnicity referents which a group of four major cross-national survey datasets collect.

Survey definitions of ethnicity are usually made in terms of categorisations based upon the boundaries of one or more alternative referents. The term categorisation is crucial, as it leads to an apparently irresolvable tension for survey researchers. For the purposes of data analysis, simple, parsimonious categorisations, with large numbers of cases in all groupings, are preferred. Candidate ethnic referents, however, have anything but those properties. They incorporate many categories with very sparse numbers of cases. And, according to social theorists, they should be cross-classified across a large number of alternative permutations.

1 Other ongoing projects in this field include the TIES project coordinated by M. Crul (see <http://www.niwi.knaw.nl/en/oi/nod/onderzoeker/PRS1258604/toon>); and the "Ethnic minority disadvantage in the labour market" project coordinated by A. Heath (see <http://www.britac.ac.uk/events/2003/031102emd-prog.html>).

2 This term is not ideal, since in other circumstances, '*ethnicity*' is taken to refer specifically to subjective cultural identity.

Table 1 Ethnic Referent Data Availability, by Country and Study

Ethnic referent:										
C	Citizen of which country			P	Parental country of birth			E	Ethnic self-identity	
B	Country of birth – which			L	Which language used			R	Religious denomination	
T	Time in this country							n	No relevant data	
<i>Letters indicate presence of data on relevant referent. Lower case letters when categories are dichotomy only – e.g., born in host country or not is 'b' rather than 'B'. Blank cells for non-coverage of country within study</i>										
	ESS	ISSP	WVS	LIS		ESS	ISSP	WVS	LIS	
<i>Australia</i>		BR	BT	BT	<i>Latvia</i>		CLR	BT		
<i>Austria</i>	CBTPLeR	R		Cb	<i>Lithu.</i>			BT		
<i>Belgium</i>	CBTPLeR			C	<i>Luxem.</i>	CBTPLeR				CT
<i>Bosnia</i>			BTR		<i>Maced.</i>			bTR		
<i>Bulgaria</i>		RE	BTR		<i>New Z.</i>		ER			
<i>Canada</i>		LER		bT	<i>Nthlds</i>	CBTPLeR	R			
<i>Croatia</i>			BTER		<i>N. Irel.</i>		ER			
<i>Cyprus</i>		ER			<i>Norway</i>	CBTPLeR	R	BTR		Bp
<i>Czech R</i>	CBTPLeR	bR		C	<i>Poland</i>	CBTPLeR	BR	bER		n
<i>Denmark</i>	CBTPLeR	BR		CT	<i>Portugal</i>	CBTPLeR	cR			
<i>Estonia</i>			BT	E	<i>Russia</i>		ER	BT		BE
<i>Finland</i>	CBTPLeR	LR	bTLR	L	<i>Serbia+M</i>			bTLR		
<i>France</i>		R		C	<i>Slovenia</i>	CBTPLeR	ER	bTIER		n
<i>Germany</i>	CBTPLeR	CR	RBTE	CB	<i>Spain</i>	CBTPLeR	R	BT		n
<i>Greece</i>	CBTPLeR				<i>Sweden</i>	CBTPLeR	cR	bTLR		CBTp
<i>Hungary</i>	CBTPLeR	ER		E	<i>Switz</i>	CBTPLeR	CLR	BT		c
<i>Ireland</i>	CBTPLeR	R		CBT	<i>UK (GB)</i>	CBTPLeR	ER	E		E
<i>Israel</i>	CBTPLeR	BR		T,R-B	<i>USA</i>		ER	BT		cE
<i>Italy</i>	CBTPLeR	R		B						

ESS : all studies from 2002.
 ISSP : all studies for 2000, except Australia, Cyprus, France, Hungary, Latvia, Poland (1999) and Italy (1998).
 WVS : Wave 3 only 1995-7 (other countries are covered by WVS in earlier waves and not W3). In some WVS countries, relevant data is nominally present, but all categories undocumented, so listed here as missing.
 LIS : uses latest available LIS study, within range 1994-2001.

Indeed, extended sociological debates on the nature of ethnic differences have generated what may be characterised a ‘paradigm of diversity’ in the field. In this model, the promotion and discussion of diversity of ethnic locations dominates contemporary discourse. At the least, the role of a great many categorical differences in defining ethnic locations is emphasised (Modood et al., 1997, 2002; the significance of boundaries may also vary over time, e.g. Aspinall, 2002). In stronger interpretations of the paradigm of diversity, many writers have argued for the inherent instability and contextuality of ethnic locations, as they are complicated by evidence of cultural hybridities (Aspinall, 2003), transnationalism (Vertovec, 2003), and/or complex social networks (Sanders, 2002). Such perspectives have lead some to the conclusion that survey variable operationalisations can offer at best a weak analytical device for representing ethnic diversity (e.g. Ahmad, 1999). Neverthe-

less, within many countries, attempts have been made to define an optimal ethnic categorisation in terms of the one or more ethnic referents thought most important³ – these constitute ‘official’ ethnic categorisations available on survey outputs.

An immediate problem for cross-national researchers is that the most relevant ethnic diversities in different countries are rarely equivalent, leading to substantial differences in the type of ethnic data which is collected by different national surveys. The cross-national comparability of categorisations which do not use the same ethnic referents is clearly questionable⁴. Indeed, many previous discussions of ethnicity measures for comparative research have ended at this point, abandoning the project (e.g. Rea et al., 1999).

Differences in data availability between countries arise both as a result of alternative perspectives in social science theories⁵, and as a consequence of nation specific institutional traditions (Lambert & Penn, 2001; Hoffmeyer-Zlotnik, 2003; Favell, 2003). The latter are often strongly politicised and rigidly enforced – see for example Favell’s (2001) description of the contrast between British and French views on the official recording of ‘racialised’ categories. Indeed, national institutions’ influence over the research data available have caused considerable difficulties to previous research programmes – illustrated by attempts at coordinating reviews of labour market situations reported in two recent European Employment Observatory Reviews (Stille, 1999; SYSDERM, 2003).

A common reaction in cross-national research methodology has been attempts to adjudicate between alternative ethnic referents in a manner which is parsimonious, cross-nationally consistent, and politically acceptable. Hoffmeyer-Zlotnik (2003) suggests a strategy for delineating the key dimensions of six ethnic referents in a manner amenable to social survey questions. This leads to the prescription of a series of differentiations that all international surveys should seek to make (2003: 276; see also Table 3). Other writers have argued the theoretical pre-eminence of certain or multiple referents as markers of ethnic differences in the cross-national context (see Wrench & Solomos, 1993; Lloyd, 1995; Smith & Blanc, 1995; Rea et al., 1999; Aspinall, 2002).

3 Typically however, consensus on a choice of categories has been elusive. See for instance critiques of the British official classification (Ballard, 1997; Aspinall, 2002, 2003; Mason, 2003) – which arise despite extended consultations with social science practitioners during their development (Sillitoe & White, 1992; Owen, 1996).

4 The comparability problem is not unique to cross-national research. Green and Owen (1995), for instance, highlight the same issues with regard to regional structures in the UK.

5 As an example, contrast Britain, where measures of religion in combination with information on subjective ethnic identity are often argued to be the most significant markers of ethnic difference (Modood et al., 2002), with the US, where the importance of diversities between immigration cohorts and their originating countries is given most prominence (Alba & Nee, 2003)

Such attempts to prioritise specific ethnic referents can be presented as ‘absolutist’ approaches to the comparative measurement of ethnic difference. A common absolutist strategy in cross-national survey research is to focus upon only a single ethnic referent. Example choices involve immigrant or citizenship status (van Tubergen et al., 2004; SYSDM, 2003); language minorities (Chiswick & Miller, 1995); or consideration of only certain distinctive ethnic groups (Model et al., 1999; Brown, 2000; Crul & Vermeulen, 2003).

Whilst such singular approaches offer an appealing conceptual clarity, they also do a disservice to the tremendous complexity of most countries’ ethnic mosaics. An alternative absolutist approach involves choosing ethnicity categorisations which are defined by boundaries across several ethnic referents. The vast sociological literature on the interdependence and complexity of multiple alternative ethnic referents suggests that such a strategy is essential: no single referent can reasonably be studied out with the context of several others (e.g. Modood et al., 1997; Alba, 2005)⁶. Selected operationalisations of singular and cross-cutting absolutist ethnic categorisations in the cross-national context are discussed below.

The search for ‘absolutist’ comparability in categorisations of ethnic difference may, however, reflect only one approach to the cross-nationally comparative measurement of ethnic difference. An alternative form of comparability may be described as ‘relativist’. Under this approach, particular ethnic categories within any country may be assigned a location within a dimension of difference which is defined relative to the structure of ethnic inequalities within the particular country⁷. Such relativist approaches to comparable measurement have been employed in other fields of cross-national survey research – for instance, measures of income, occupation or education standardised according to national averages (cf. Hoffmeyer-Zlotnik & Wolf, 2003). Relativist measures have not been widely considered in the context of measures of ethnicity, although recent papers by Lambert and Penn (2001), and Alba (2005), have taken this stance⁸. We argue below that relativist measures of ethnic difference have several attractions to cross-national survey researchers.

6 This same point may be phrased in terms of ‘omitted variable bias’ or ‘spuriousness’. The exclusion of information on some relevant referents could lead to an impartial and misspecified account of the influence of others. Chiswick and Miller (1995), for instance, illustrate that information on language ability (referent 1) can be a highly relevant explanatory variable in studies of the economic impact of immigrant status (referent 2).

7 Hoffmeyer-Zlotnik & Wolf (2003) note how ‘functional equivalence’ is required in order to make cross-nationally comparative statements. This equivalence requires that measurement from diverse countries always refers to the same dimension of difference. Both ‘absolutist’ and ‘relativist’ categorisations can be presented as functionally equivalent in this regard.

8 An earlier relativist proposal, applied specifically to Britain, was made by Prandy (1979). van Tubergen’s (2004) analysis may also be characterised as a relativist approach, as it incorporates measures of immigration background which refer to the relative characteristics of the sending and receiving communities.

3 Measurement of Ethnicity within Contemporary Surveys

The briefest examination of existing survey resources quickly reveals the difficulties of obtaining cross-nationally harmonised information on ethnic differences. Tables 1 and 2 illustrate a number of patterns in the data resources of four cross-nationally harmonised survey data collections. These are the European Social Survey (ESS, see Jowell, 2003 and www.europeansocialsurvey.org/; the ESS data is accessed via the Norwegian Social Science Data Services); the International Social Science Project datasets (ISSP, see Braun & Uher, 2003 and www.issp.org ; the ISSP data was accessed via the UK Data Archive at the University of Essex); the World Values Survey datasets (WVS, see Inglehart, 2000 and www.worldvaluessurvey.org/; the WVS data was accessed via the UK Data Archive at the University of Essex); and the studies of the Luxembourg Income Study (LIS, see an earlier review as Lambert & Penn, 2001, or www.lisproject.org; the LIS may be accessed direct via the project webpages). The survey collections differ considerably in their data collection strategies. The ESS attempts to follow the highest standards of 'pre-harmonisation' of questions before going to field, resulting in high levels of consistency between questions across different countries. Both the WVS and ISSP similarly use pre-harmonisation techniques, although more flexibility between countries and time periods is built into their designs. Lastly the LIS surveys are entirely 'ex-post' harmonised, one consequence being more examples of incomparable questions between countries, as well as higher risks of coding and translation errors (e.g. van Deth, 2003)⁹.

Tables 1 and 2 can be summarised by two concise points¹⁰. The first (see Table 1) is that most of the surveys, and particularly the ESS, have nominally a very good collection of ethnicity related data, covering several alternative ethnic referents. The catch, however (see Table 2), is that the uses to which that source data could be used are highly constrained, given the sparsity of the responses involved.

9 Many other surveys may have been considered, though these four studies illustrate very typical properties of the medium-scale national sample surveys which are perhaps most widely used in cross-national survey comparisons.

10 Fuller elaborations of these tables, breaking down many of the components by particular countries, may be obtained on request from the author (also www.staff.stir.ac.uk/paul.lambert/downloads.html).

Table 2 'Wealth' of Ethnicity Data on 4 Survey Collections*(Ethnic referent codes as Table 1)*

#Cat	Number of categories in original data								
#NSC	Number of non-sparse categories (more than 50 cases, absolute value)								
Skew	Skewness: Percent of valid cases in the largest category								
%m	Percent of cases with missing data								
	#Cat	#NSC	Skew	%m		#Cat	#NSC	Skew	%m
ESS 2002: pooled data, 21 countries					ISSP 2000: pooled data, 28 countries				
	Country of citizenship				C	10	1	95	0
C	118	5	96	0	B	20	2	80	2
	Country of birth				L	4	2	60	0
B	158	12	90	0	E	10	2	80	20
	Time living in country (categories)				R	8	3	50	2
T	6	5	90	0	ISSP: Estimates are typical values per country				
	Language spoken at home				<i>Typical n = 1000 cases / country</i>				
L	102	9	95	1	WVS 1995/7: pooled data, 20 countries				
	Whether in a minority ethnic group				B	100	12	91	2
E	2	2	94	2	T	7	7	92	21
	Parents national origins (continent of parents' birth / mixed parentage)				L	100	12	95	9
P	7	6	84	0	E	200	10	90	14
	Religious denom. (current or past)				R	14	11	38	4
R	9	9	40	1	<i>WVS 1995-7: Estimates are absolute values for merged dataset. Typical n = 1000 cases/country</i>				
<i>ESS 2002: Estimates are absolute values for merged dataset. Typical n = 2100 cases / country.</i>					LIS c1995: pooled data, 24 countries				
ESS Data – illustrative countries:					C	50	2	90	1
	UK (GB), n=2052				B	50	4	90	0
C	28	1	97	0	T	m	m	90	1
B	57	1	91	0	L	2	2	94	1
T	6	2	91	0	E	10	3	95	2
L	31	1	96	0	P	7	5	90	0
E	2	1	94	0	<i>LIS: Estimates are typical values per country</i>				
P	7	3	85	1	<i>Typical n = 50000 cases / country</i>				
R	9	4	45	0					
	Germany, n=2919								
C	43	1	96	0					
B	51	1	93	0					
T	6	2	93	0					
L	20	1	96	0					
E	2	2	96	0					
P	6	3	86	0					
R	9	5	38	1					

However, closer inspection of the data from any particular study quickly reveals the superficiality of many of the apparently complete records. Table 2 summarises the ‘wealth’ of the ethnic referent data across the surveys¹¹. Almost all of the differentiations recorded are categorical in nature, and the first column [#Cat] indicates the number of unique categories measured by the relevant variables. The second and third columns then indicate the sparsity and skew of the relevant distributions. Skewness, as the proportion of cases clustering into the largest category [Skew], is an indicator of how much variation is likely to be usefully analysed – a highly skewed variable can offer little differentiating information between cases, especially if the overall sample size is relatively small. The sparsity measure [#NSC] checks the absolute number of cases in relevant ethnicity categories, listing the number of categories with more than 50 people representing them. This again indicates how much analysis can realistically be undertaken on the variables – the absolute number of cases is ultimately more important than skew in this regard. Lastly, the fourth column [%m] shows the number of missing cases for each relevant variable – in some literatures it is generally expected that questions relating to ethnicity will be characterised by high levels of missing data, though this is not generally borne out across the range of datasets shown.

Table 2 summarises these patterns of data ‘wealth’ for the pooled cross-national samples (ie, with data from different countries combined). Although crude¹², this serves to illustrate the key features of the survey data. This is of a clear pattern of ‘impoverished’ data resources on ethnicity for the majority of countries. Although missing data is minimal, the data distributions are highly sparse and skewed, so that there are in practice likely to be very few circumstances where the divisions will sustain an informative analysis of ethnicity.

Compounding these basic distributional weaknesses, closer inspection of many of the datasets reveals that they often fail to measure some of the more theoretically appealing features of ethnic difference. For instance, few studies have extended data on subjective ethnic identity divisions (the ESS for example employs a dichotomous record of whether or not the subject identifies with any minority group). Few countries have substantial data on parental place birth (for the ESS, the data only differentiates 5 alternate continents). And the apparent ‘wealth’ of responses to questions on religion turns out in most cases to hinge on substantively less interesting divisions between categories of ‘no religion’ and one or more major Christian churches.

11 We use the term ‘wealth’ in reference to survey analysts’ common descriptions of the ‘richness of the data’: wealthy data contains high variability across cases on the relevant variables, and thus ‘relationships between variables have the maximum chance to show up’ (Punch, 2003: 38).

12 See Footnote 10

The relatively small samples of the ESS, ISSP and WVS offer further problems of data wealth. However it can be noted that within each of these studies, some specific countries do exhibit greater variation across ethnicity measures in a 'wealthier' way. These three studies are also attractive to researchers because the elements of pre-harmonisation in their design make data access and documentation relatively easy¹³. Although the LIS data studies have similar levels of skew in their relevant distributions, it can be noted that their number of cases is often much larger, thus meaning that sparsity of representation of categories can be less of a problem. The LIS data may also be noted for the wide range of ethnic referents measured between countries. This reflects greater national specific variations in data collection. This contributes both to conceptual confusion, but also to increasing the degree to which the LIS resources engage with ethnic referents of national specific research concern.

The picture of ethnicity information available on these four groups of surveys is messy and problematic. It should be remembered, moreover, that harmonised survey resources like these represent the stronger examples of comparative survey resources. Other researchers have reported greater difficulties attempting to conduct comparative analysis involving ethnicity on survey data collections which have not been subjected to the same levels of harmonisation and documentation (for instance, Stille, 1999).

4 Absolutist Solutions to Cross-National Research on Ethnicity

As Tables 1 and 2 have illustrated, the type of ethnic referent data collected by national surveys is inconsistent, and its analysis problematic due to the sparse numbers of cases representing many of the minority categories. An 'absolutist' perspective to the harmonisation of ethnicity data, therefore, suggests a pragmatic strategy, of choosing categories which may be both readily operationalised in most survey collections, and which will have modest numbers of cases in their minority categorisations. We list below six 'absolutist' categorical definitions of ethnicity differences which generate at least moderately 'wealthy' data structures for the survey datasets under study:

- (1) [IMM] **Immigrant status:** a dichotomy indicating whether or not a case was born in the current country. This measure is widely used in previous research (though many studies record the age at entry to the host country). However it conveys limited information about ethnic differences, as it fails to make internal differentiations

13 A drawback in working with the WVS data on ethnicity is that the level of documentation in the centrally distributed files can be somewhat limited: readers must cross-check between multiple documents to obtain pithy descriptions of category labels, and, in a few example countries, apparent ethnic group data has no accessible documentation pertaining to it (thus excluded from Tables 1 and 2).

between different immigrant backgrounds, and it cannot recognise non-immigrant minorities, or even the children or grandchildren of immigrant ancestors (e.g. Banton, 1997). It is also flawed because of its popular conflation with concepts of citizenship – Hoffmeyer-Zlotnik (2003) emphasises how the two measures are analytically distinct).

- (2) **[LAN] Minority language use:** a dichotomy indicating whether or not a case generally speaks a language other than an official majority language of the host country. This measure has considerable sociological significance, as the analysis of ethnic differences is increasingly concerned with information on language use (e.g. Portes & Rumbaut, 2001; Alba & Nee, 2003). However this measure has several flaws: there is a lack of cross-national consistency to the measurement of ‘minority’ languages, and a subjectivity to individuals’ reports of usage. It also ignores ethnic minorities who do not use minority languages.
- (3) **[VIS] Visible minority group status:** a dichotomy indicating whether or not a case belongs to a minority group on the grounds of any overt ethnic group formation – such as racialised visibility, subjective ethnic group identity, or participation in a ‘visible’ minority religion¹⁴. This measure highlights the ethnicity referents which are most popular with leading sociological thinking in the field (cf. Modood, 2002); however, it conceals considerable internal heterogeneity.
- (4) **[MIN] Any minority group membership:** a dichotomy which extends the VIS categorisation to highlight any ethnic minority identity from any relevant referent – citizenship, country of birth, parental national origins, language use, and visible minority group status. Again, this catholic measure masks diversity within its categories. It is also unlikely that the same differentiations will contribute to the same categories in different countries.
- (5) **[CON] National-specific scheme:** a multiple categorical scheme chosen from the available schemas favoured by national literatures. This has clear substantive attractions, but is problematic for comparative purposes.
- (6) **[EC9] Comparative ethnicity 9 category measure:** a multiple categorical scheme advocated here as an attempt at incorporating as many ethnicity differentiations (cf. Hoffmeyer-Zlotnik, 2003) as reasonably possible. It uses information from five dichotomous measures, IMM, LAN, VIS, and indicators of citizenship and of whether or not either parent was born outside the country. That information is cross-classified to form the categories listed below, where ‘minority group’ corresponds to either VIS=1, or having non-host citizenship:

14 This religion criteria used here reflects the perspective (and approximation), advocated for instance by Modood et al. (2002) and Brown (2000), that only certain minority religions are most important to social stratification outcomes in the contemporary societies under study: here defined as recording a South Asian or Islamic religious identity.

- 1 CCNN – Self and parents born in country, no minority group or language
- 2 CCMN – Self and parents born in country, minority group but not language
- 3 CCL – Self and parents born in country, minority language
- 4 CPNN – Self born in country, parents not, no minority group or language
- 5 CPMN – Self born in country, parents not, minority group but not language
- 6 CPL – Self born in country, parents abroad, minority language
- 7 FNN – Self born abroad, no minority group or language
- 8 FMN – Self born abroad, minority group but not language
- 9 FL – Self born abroad, minority language

Table 3 summarises the distribution of cases to these categorical schemes from some of the survey collections considered. Its distributions are indicative of a number of important points. Firstly, the four dichotomous measures – whilst distinctly unsatisfactory as a theoretical categorisation – have more attractive data distributions. On the other hand, the national specific categorisations remain highly problematic – the skewness and sparsity of their minority categorisations is pronounced, whilst there is inconsistency in the number of categories delineated between countries. Lastly, the proposed ‘EC9’ scheme appears to be reasonably appropriate. It largely avoids sparsity problems (with the exception of the CCL and CPL categories, minority groups with regard to language but not other criteria), and is in most circumstances capable of being derived for most countries of interest (with some exceptions).

The value of these alternative categorisations can also be assessed by empirical judgements on the strength of association between ethnicity indicators and other social measures. When we analyse, for example, indicators of social stratification advantage, we see clear patterns of association with these six ethnicity measures. Importantly, the four dichotomies (particularly the ‘MIN’ categorisation), prove relatively strong, accounting for the large bulk of outcome variation that can be attributed to further ethnicity differentiations. Also of interest is the observation that slightly more variation is typically associated with the EC9 measure (which has no national specific input), than with the CON measure. However, other analyses of these ethnicity indicators leads to less favourable conclusions. We frequently observe non-significant main, and interaction, effects associated with dummy variable indicators of ethnic minority groups (variously defined), when used as predictors of social stratification advantage in human capital style regression models. On theoretical grounds we have strong reasons for expecting such associations to exist – it can be assumed that the small number of cases in many minority groups is making evidence of their effects difficult to confirm.

Table 3 ‘Absolutist’ Ethnicity Measures: Properties of Alternative Categorical Differentiations (n’s representing ethnic categories, ESS/WVS/ISSP data)

	ESS	WVS		ESS	WVS	ISSP	
						UK	Germ.
IMM (<i>if born in host country</i>)			CON (<i>national specific ethnicity categories</i>)				
0	36839	21736	1			919	1417
1	3935	2129	2	n/a	n/a	12	20
			3			7	10
LAN (<i>if speak minority language</i>)			4			9	15
0	36383	20311	5			5	37
1	4336	3516	6			2	
			7			1	
VIS (<i>if a ‘visible’ minority group</i>)			8			3	
0	36261	16655	9			13	
1	3655	5236					
			EC9 (<i>minority categories by cross-classification of referents</i>)				
MIN (<i>if minority by any referent</i>)			1 – CCNN	29176	1405	1642	2432
0	29176	14015	2 – CCMN	832	2048	60	47
1	10716	6886	3 – CCL	547	1783	1	1
			4 – CPNN	2051		92	161
cf.: Hoffmeyer-Zlotnik (2003: 276)			5 – CPNM	850		34	19
1 – Legal status (citizenship)			6 – CPL	122		5	11
2 – Country of birth			7 – FNN	1205	597	63	59
3 – Visibility			8 – FMN	1216	292	68	64
4 – Mother tongue			9 – FL	1115	847	50	69
5 – Cultural customs / religion			% missing	9.2	25.5	1.8	1.9
6 – Self-perception							

This evaluation of ‘absolutist’ representations of ethnic differences helps illustrates several issues. Firstly, a number of dichotomous measures of concepts are readily operationalised in cross-national surveys and exhibit discernible associations with other relevant factors; they may also be easily understood and communicated. They are, however, compromised by substantial internal heterogeneity, which may mask otherwise significant ethnic differences. Alternatively, national specific measures (‘CON’ variables) are seriously constrained, by a lack of conceptual comparability between countries; by the lack of availability of the relevant data for many surveys; and by the sparse representation of many contributing categories. Lastly, the ‘EC9’ categorical representation has both conceptual attractions (as a consistent measure between countries) and weaknesses (its possible disengagement from national-specific concerns over ethnicity differences). As a practical measure EC9 has intermediate properties. Its association with other variables is particularly strong, and in most circumstances there is a moderate spread of cases between different categories. However, categories of the measure still tend towards sparsity in certain contexts, whilst many of its categories continue to conflate sociologically signifi-

cant ethnicity differences. It may thus be argued that with harmonised social surveys, it remains difficult to devise and analyse satisfactory categorical ethnicity classifications which apply to national populations in a meaningful way.

5 Relativist Solutions to Cross-National Research on Ethnicity

A ‘relativist’ approach to recording ethnic differences involves two stages. The first is to record ethnic locations, as mapped by information on multiple possible ethnic referents. The second is to analyse that location, not simply as a qualitative category, but as a position within a dimension of relative ethnic difference. This is often achieved mathematically, by assigning metric scores to different categorical locations (for earlier developments of this argument, see Lambert & Penn, 2001).

Table 4 gives an illustrative selection of such derived ethnic category scores. The columns of Table 4 show the results from a series of Stereotyped Ordered Regression (SOR) models (Anderson, 1984). Here, ethnic category scores are derived as a function of average differences between ethnic categories according to a variety of predictor variables¹⁵. Metric scores derived by SOR models for a selection of EC9 and CON categories are shown: the values represent a dimension of difference between categories, as they are associated with a regression format model summarising demographic and economic effects. The precise interpretation of the scores hinges upon the parameter coefficients of the regression models on which they are derived (not shown), but the important point is that the SOR scores represent average relative differences between individuals from different ethnic categories in a country along a dimension of measured social and economic circumstances.

In this case, a scaling is thus derived which reflects *relative* ethnic difference within each country. In practice, such SOR scores could be derived from national specific information sources (with greater coverage of cases), then mapped back to the sparser categories identified in sample surveys. In principle, SOR scores could be estimated for the dozens of different categories which emerge when cross-classifying multiple ethnic referents. The subsequent metric treatment of these categories can then be exploited to avoid the problems of undertaking categorical data analyses techniques with sparsely represented groups.

15 The SOR models were derived using macros following Hendrickx (2000); for discussion of the specification of these models for this context, see Lambert and Penn (2001).

Table 4 'Relativist' Ethnicity Measures: 'SOR' Score Derivations for Selected ESS Samples

SOR regression: $[Ethnicity] = gender + age + education + marital\ status + employment\ status + employment\ advantage + interactions$

	North-West Europe sample, n=20899		UK only, n=1893		Germany, n=2503	
SOR scores for EC9	EC9	EC9	EC9	CON_U	EC9	CON_G
1. CCNN	-239	-530	-253		-332	
2. CCMN	-220	-144	-152		-369	
3. CCL	-380†	-368†	-	-571	293†	-504
4. CPNN	-118	-22	-226	309†	-328	826†
5. CPMN	233	192	-143	643†	461†	-235
6. CPL	634	604	717†	22	324†	-87
7. FNN	-330	-203	484	-404	-387	
8. FMN	5	171	-241		33	
9. FL	414	300	-186		305	

† : Less than 20 cases in the SOR category
 CON_U: 1=White; 2=Black-Caribbean; 3=Black-African; 4=Asian; 5=Other.
 CON_G: 1=German citizen; 2=Turkish citizen; 3=Eastern European citizen; 4=Other citizen.

There are of course many other ways in which ethnic categories could reasonably be assigned score values. For instance this could be through using other summarising functions which have perhaps simpler and more communicable interpretations (say the average employment advantage score associated with each group), or perhaps by *fiat*, allowing sociological thinkers an opportunity to make judgements reflecting a wider range of considerations concerning ethnic differences. However derived, category scoring offers an opportunity to analyse ethnicity differences in hierarchical manners, interpreted as the typical effects of ethnicity differences as they operate through a given structure of inequalities¹⁶.

There are several attractions to this strategy. The metric representation can bypass many data analytical difficulties associated with the low representation of minority group categories. The metric representation can itself prove a highly revealing communication on the nature of ethnic differences within any given country. A relativist metric can chime more closely with recent sociological theorising on the complexity, diversity, dynamism and contextuality of ethnic differences, as its flexibility can allow the mapping of multiple ethnic locations defined in terms of numerous cross-cutting referents. Lastly, it can be

16 It is important to note that metric variable representations should not be conflated with uni-dimensionality – multiple dimensions of differences can be incorporated (Lambert & Penn, 2001).

argued that relative measures of ethnic differences have a form of cross-national comparability so long as the underlying structure of ethnic differences is estimated in a similar way across countries – in the same fashion, for example, as a measure of income that is standardised around national averages.

The potential value of such ethnic category scoring is that it may parsimoniously summarise the key elements of ethnic differences that affect the population under study. Empirical analyses have so far only partially supported this view. For instance, the derived scores presented in Table 4 exhibit only modest patterns of association with other variables which are expected to relate to ethnic differences. Lambert and Penn (2001) similarly found only a few circumstances where SOR score effects altered the estimates from categorically based analytical regression models. Nevertheless, there is a strong possibility that a more powerful relativist scoring framework could be uncovered, and thus that the assignment of ethnic category scores could prove a significant benefit to cross-national analysis.

6 Conclusions

This review has identified three significant problems in the analysis of ethnicity in cross-national studies. Firstly, it has shown that existing surveys resources are haphazard in their measurement of ethnic referents between different countries and time periods. Certain more consistent choices in data collection strategy could be employed (e.g. Hoffmeyer-Zlotnik, 2003), though to date, the European Social Survey is the only major harmonised collection which has come close to achieving these standards. Secondly, we have seen that regardless of the range of relevant questions covered in surveys, in almost all instances, harmonised survey collections suffer severe problems of data ‘wealth’ with regard to ethnicity information. This is primarily a function of the sparse representation of the minority groups. Thirdly, this review has also demonstrated severe theoretical critiques of the cross-national analysis of ethnicity effects. These focus upon the apparent inability of harmonised categorisations to reflect the full and appropriate ‘ethnic mosaic’ of the societies under study. However, it is argued that despite such problems, we should not shy away from the analysis of ethnicity in a comparative perspective, and this review has sought to outline reasonable solutions from two alternative perspectives.

A ‘relativist’ approach to the harmonisation of ethnicity data involves categorisations, from whichever ethnic referent is available, being mapped onto one or more dimensions of ethnic inequality, and given numerical indicator scores for their relative positions. This strategy is cognitively challenging, as it requires the researcher to communicate abstract ideas on relative positions within a hypothesised social structure. It is also empirically

ambiguous, as it is only in selected circumstances that relativist measures of ethnic difference appear to exhibit greater criterion validity than absolutist measures. However, relativist measures have considerable theoretical appeal, since they are far more amenable to cross-national (and temporal) variations in the type of ethnic data collected, and offer an abstract form of functional equivalence which exceeds those of absolutist schemes.

'Absolutist' approaches to harmonisation, which involve ensuring that the same concepts are used to differentiate categories in different countries, are more easily communicated, and may be expected to continue to be widely used. This strategy can lead to crude measures of ethnic difference, perhaps involving only one ethnic referent or the merger of a number of loosely related minority group categories. However it can be argued that certain carefully applied absolutist categorisations can be more satisfactory. The 'EC9' categorisation illustrated above, for instance, has reasonably strong analytical properties and is apparently appropriate for the majority of countries considered in this review.

References

- Ahmad, W. 1999. "Ethnic Statistics: Better than nothing or worse than nothing?" Pp. 124-131 in *Statistics in Society: The Arithmetic of Politics* edited by Dorling, D., and S. Simpson. London: Arnold.
- Alba, R. 2005. "Bright vs. blurred boundaries: Second-generation assimilation and exclusion in France, Germany, and the United States." *Ethnic and Racial Studies* 28:20-49.
- Alba, R., and V. Nee. 2003. *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Harvard: Harvard University Press.
- Allen, S., and M. Macey. 1990. "Race and Ethnicity in the European Context." *British Journal of Sociology* 41:375-393.
- Anderson, J. A. 1984. "Regression and Ordered Categorical Variables." *Journal of the Royal Statistical Society Series B* 46:1-30.
- Anthias, F. 2001. "The concept of 'social division' and theorizing social stratification: Looking at ethnicity and class." *Sociology* 35:835-854.
- Aspinall, P. J. 2002. "Collective terminology to describe the minority ethnic population: The persistence of confusion and ambiguity in usage." *Sociology* 36:803-816.
- Aspinall, P. J. 2003. "The conceptualisation and categorisation of mixed race/ethnicity in Britain and North America: Identity options and the role of the state." *International Journal of Intercultural Relations* 27:269-296.
- Ballard, R. 1997. "The construction of a conceptual vision : Ethnic groups and the 1991 UK census." *Ethnic and Racial Studies* 20:182-194.
- Banton, M. 1997. *Ethnic and Racial Consciousness*. London: Longman.

- Bonifazi, C., and S. Strozza, (Eds.). 2003. *Integration of migrants in Europe: data sources and measurement in old and new receiving countries*. Special issue of *International Journal of Migration Studies* 40(152).
- Braun, M., and R. Uher. 2003. "The ISSP and its Approach to Background Variables." Pp. 33-48 in *Advances in Cross-National Comparison*, edited by Hoffmeyer-Zlotnik, J. H. P., and C. Wolf. New York: Kluwer.
- Brown, M. S. 2000. "Quantifying the Muslim population in Europe: conceptual and data issues." *International Journal of Social Research Methodology* 3:87-102.
- Castles, S. 1993. "Migrants and Minorities in Europe. Perspectives for the 1990s: Eleven Hypotheses". Pp. 17-34 in *Racism and Migration in Western Europe*, edited by Wrench, J., and J. Solomos. Oxford: Berg.
- Chiswick, B. R., and P. W. Miller. 1995. "The Endogeneity between Language and Earnings – International Analyses." *Journal of Labor Economics* 13:246-288.
- Crul, M., and H. Vermeulen. 2003. "The Second Generation in Europe." *International Migration Review* 37:965-986.
- Evans, G., and A. Need. 2002. "Explaining ethnic polarization over attitudes towards minority rights in Eastern Europe." *Social Science Research* 31:653-680.
- Favell, A. 2001. *Philosophies of Integration: Immigration and the Idea of Citizenship in France*. 2nd Edition. Palgrave: Macmillan.
- Favell, A. 2003. "Integration Nations: The Nation-State and Research on Immigrants in Western Europe." *Comparative Social Research* 22:13-42.
- Green, A. E., and D. Owen. 1995. "Ethnic minority groups in regional and local labour markets in Britain: A review of data sources and associated issues." *Regional Studies* 29:729-735.
- Harkness, J. A., F. J. R. van de Vijver, and P. P. Mohler. 2003. *Cross-Cultural Survey Methods*. New York: Wiley.
- Heckmann, F., R. D. Penn, and D. Schnapper. 2001. *Effectiveness of National Integration Strategies Towards Second Generation Migrant Youth in a Comparative Perspective – EFFNATIS*. Bamberg: European Forum for Migration Studies, University of Bamberg.
- Hendrickx, J. 2000. "Special Restrictions in multinomial logistic regression." *Stata Technical Bulletin* 56:18-26.
- Hoffmeyer-Zlotnik, J. H. P. 2003. "How to Measure Race and Ethnicity" Pp. 267-277 in *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*, edited by Hoffmeyer-Zlotnik, J. H. P., and C. Wolf. New York: Kluwer Academic.
- Hoffmeyer-Zlotnik, J. H. P., and C. Wolf (Eds.). 2003. *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*. New York: Kluwer Academic.
- Hutchinson, J., and A. D. Smith (Eds.). 1996. *Ethnicity*. Oxford: Oxford University Press.

- Inglehart, R. 2003. *World Values Surveys and European Values Surveys 1981-4, 1990-3, 1995-7* [Computer file]. Ann Arbor, MI : Institute for Social Research [Producer]; Ann Arbor, MI : Inter-university Consortium for Political and Social Research [Distributor]. 26.7.02.
- Jacobs, D., and J. Tillie. (Eds.). 2004. "Social Capital and Political Integration of Migrants." *Journal of Ethnic and Migration Studies* 30:419-427.
- Jowell, R. 2003. *European Social Survey 2002/2003: Technical Report*. London: Centre for Comparative Social Surveys, City University.
- Lambert, P. S., and R. D. Penn. 2001. *SOR models and Ethnicity data in LIS and LES : Country by Country Report*. Syracuse University, Syracuse, New York 13244-1020: Luxembourg Employment Study Paper No. 21, Maxwell School of Citizenship and Public Affairs. (<http://www.lisproject.org/wpapers.htm>).
- Lloyd, C. 1995. "International Comparisons in the Field of Ethnic Relations." Pp. 31-46 in *Racism, Ethnicity and Politics in Contemporary Europe*, edited by Hargreaves, A. G., and J. Leaman. Aldershot: Edward Elgar.
- Model, S., G. Fisher, and R. Silberman. 1999. "Black Caribbeans in comparative perspective." *Journal of Ethnic and Migration Studies* 25:187-212.
- Modood, T., R. Berthoud, J. Lakey, J. Y. Nazroo, P. Smith, S. Virdee, and S. Beishon. 1997. *Ethnic Minorities in Britain: Diversity and Disadvantage*. London: Policy Studies Institute.
- Modood, T., R. Berthoud, and J. Nazroo. 2002. "'Race', racism and ethnicity: A response to Ken Smith". *Sociology* 36:419-427.
- Owen, D. 1996. *Towards 2001: Ethnic Minorities and the Census*. Warwick: Centre for Research in Ethnic Relations, University of Warwick.
- Portes, A., and R. G. Rumbaut. 2001. *Legacies: The Story of the Immigrant Second Generation*. Berkeley and Los Angeles: University of California Press.
- Prandy, K. 1979. "Ethnic discrimination in employment and housing." *Ethnic and Racial Studies* 2:66-79.
- Punch, K. F. 2003. *Survey Research: The Basics*. London: Sage.
- Rea, A., J. Wrench, and N. Ouali. 1999. "Discrimination and Diversity." In *Migrants, ethnic minorities and the labour market : Integration and Exclusion in Europe*, edited by Wrench, J., A. Rea, and N. Ouali. London: Palgrave.
- Sanders, J. M. 2002. "Ethnic boundaries and identity in plural societies." *Annual Review of Sociology* 28:327-357.
- Sillitoe, K., and P. H. White. 1992. "Ethnic Group and the British Census : the search for a question." *Journal of the Royal Statistical Society, Series A: Statistics in Society* 155:141-163.

-
- Smith, D. M., and M. Blanc. 1995. "Some Comparative Aspects of Ethnicity and Citizenship in the European Union." Pp. 70-92 in *Migration, Citizenship and Ethno-National Identities in the European Union*, edited by Martiniello, M. Aldershot: Avebury.
- Stille, F. 1999. *Ethnic Minorities and Immigrant Groups on the Labour Market : Introduction*. SYSDÉM: Employment Observatory Trends 32.
- SYSDÉM. 2003. "Dealing with the Challenge of Immigration." *European Employment Observatory Review*. Autumn, 27-112.
- van Deth, J. W. 2003. "Using Published Survey Data." Pp. 329-346 in *Cross-Cultural Survey Methods*, edited by Harkness, J. A., F. J. R. van de Vijver, and P. Ph. Mohler. New York: Wiley.
- van Tubergen, F., I. Maas, and H. Flap. 2004. "The Economic Incorporation of Immigrants in 18 Western Societies: Origin, Destination and Community Effects." *American Sociological Review* 69:704-727.
- Vertovec, S. 2003. "Migration and other modes of transnationalism: Towards conceptual cross-fertilization." *International Migration Review* 37:641-665.
- Wrench, J., and J. Solomos. 1993. *Racism and Migration in Western Europe*. Oxford: Berg.