

## Some old and new measures using educational qualifications

*Presentation to the expert workshop:*

**'Data on educational qualifications in social survey research'**  
**Paul Lambert, University of Stirling, 31 Aug 2010**

*GEEDE is a project within 'DAMES', an ESRC research Node in 'Data Management through e-Social Science', [www.dames.org.uk](http://www.dames.org.uk)*

1. Handling complex categorical data

2. Some old measures – recording and using educational data

3. New measures: Issues of functional form

- Deciding between big and small categories
- Why scales are better than big categories!

4. New measures: A CAMSIS approach for scaling educational qualifications

# 1) Handling complex categorical data

- Arguably the most common social survey datum
  - o Basis of most social research reports/analyses/comparisons
- It's rich and complex
  - o We're often interested in very fine levels of detail / difference
  - o We usually recode categories in some way for analysis

➤ ...how categorical data is *managed* is of great consequence to the results of analysis...

*Choices about recoding, boundaries, contrasts made [e.g. RAE analysis: <http://www.dames.org.uk/rae2008/>]*



Q. 129 Which of the following best describes how you would describe yourself?	Freq.		
1. British	137	31. English, Indian-British, Asian-British	1
2. English	79	32. Pakistani-British, Asian-British (6)	6
3. Pakistani	25	33. Pakistani, Asian (3,9)	3
4. Indian	10	34. Kashmiri	4
5. Bangladeshi	10	35. Pakistani-Moslem living in Great Br	1
6. Pakistani-British	73	36. Pakistani-British, Black-British, A	1
7. Indian-British	34	37. Pakistani-British, Asian, Black-Bri	1
8. Bangladeshi-British	22	38. Pakistani, Black (3,12)	1
9. Asian	6	39. Pakistani-born British	1
10. White	7	40. Pakistani-British, Asian-British (6)	2
11. European	6	41. Black, Asian-British (12,15)	2
13. White-British	60	42. Pakistani-British, Asian (6,9)	6
15. Asian-British	65	43. English, Indian-British, Asian-Brit	1
16. Cosmopolitan	7	44. English, Pakistani-British, Indian-	2
17. Moslem	12	45. Neapolitan	1
18. Indian, Asian (4,9)	2	46. Indian, Asian, Asian-British (4,9,1	5
19. British Moslem	6	47. Indian-British, Asian (7,9)	5
21. Indian & British (1,4)	1	48. English, Indian-British, Black, Asi	2
22. Indian-British, Asian-British (7,15)	9	49. English, Indian-British (2,7)	2
23. English & White (2,10)	54	50. English, Asian-British (2,15)	2
24. English, White-British (2,13)	41	51. Pakistani-British, Indian-British (	1
25. White-British & Cosmopolitan (13,16)	5	52. English, Indian, Asian-British (2,4	1
26. English & White-British (2,13)	8	53. British, English (1,2)	9
27. White-Italian	1	54. British, White (1,10)	2
28. European and White-British (11,13)	7	55. Scottish	1
29. English, European and White-British	20	56. English, European (2,11)	2
30. English, White, Cosmopolitan (2,10,	1	57. Individual	2
		58. Individual	2
		59. English, European, White-British, C	3
		60. British, White, European, Cosmopoli	1
		61. English, Pakistani-British (2,6)	2
		62. Pakistani-British, Black, Asian-Bri	2
		63. Bangladeshi, Asian, Black (5,9,12)	1
		64. British, European (1,11)	2
		65. Lancastrian	3
		66. Humanoid	1
		67. Pakistani, Asian-British (3,15)	1
		68. Pakistani, Asian-British, Moslem, F	1
		69. English, White, Irish	1
		70. English, White, European (2,10,11)	3
		71. Irish-English	1
		72. English, White-British, Cosmopolita	2
		73. British with Baltic-Slav origins	1
		77. English, Pakistani (2,3)	3
		78. Pakistani-British, Asian, Moslem	1
		79. English, Indian-British, White, Asi	1
		80. Human being	2
		82. English, Pakistani-British, Asian (	1
		83. Pakistani-British, Italian	1
		85. European, White-British, Irish	1
		86. English, Indian (2,4)	4
		87. Indian-British, Black-British (7,14	1
		88. Individual	1

EFFNATIS sample (1999):  
Subjective ethnic identity




MUSICID	1. Autoch	2. Pakist	3. Indian	4. Bangla	Total	
1. Pop	107	32	20	4	163	
2. Chart	27	2	0	0	29	
3. Dance	94	5	9	1	109	
4. Soul	8	31	0	0	39	
5. R'n'B	19	16	0	0	35	
6. Hip Hop	5	6	0	0	11	
7. Asian Music	0	11	0	0	11	
8. Trance	0	0	0	0	0	
9. House	0	0	0	0	0	
10. Techno	18	0	0	0	18	
11. Garage	0	2	0	0	2	
12. Hardcore	3	0	0	0	3	
13. Acid Jazz	1	0	0	0	1	
14. Rap	4	11	0	0	15	
15. Indie / Britpop	46	4	0	0	50	
16. Rock / Heavy Metal	32	2	0	0	34	
17. Motown	1	0	0	0	1	
18. Folk	1	0	0	0	1	
20. Jazz	2	1	0	0	3	
21. Classical	4	2	0	0	6	
22. Sixties	3	0	0	0	3	
23. Seventies	2	0	0	0	2	
24. Eighties	1	0	0	0	1	
26. Grunge	3	0	0	0	3	
27. Easy Listening	0	3	0	0	3	
28. Reggae	2	0	2	1	5	
30. Country	3	0	0	0	3	
31. Punk	0	0	0	0	0	
35. Indian	0	11	3	1	15	
36. Bangra	0	5	2	0	7	
37. Jungle	1	0	0	0	1	
38. Asian Pop Music	0	2	0	0	2	
40. All	19	5	2	0	26	
41. Drum 'n' Bass	3	0	0	0	3	
43. Alternative	2	0	0	0	2	
44. Soft Rock	2	1	0	0	3	
45. Rock 'n' Roll	1	0	0	0	1	
46. Love songs / ball	1	0	0	0	1	
47. Blues	1	0	0	0	1	
48. Swing	0	1	0	0	1	
50. Break Beat	1	1	0	0	2	
51. Goth	1	0	0	0	1	
60. Dance Floor	2	0	0	0	2	
63. English	0	2	1	0	3	
64. Classical Hindi	0	1	0	0	1	
65. Hindi	0	1	2	0	3	
68. Urdu	0	2	0	0	2	
72. Lo-fi	1	0	0	0	1	
75. Ambient	1	0	0	0	1	
77. Bollywood	0	0	1	0	1	
78. Dramatic	0	0	4	0	4	
79. Islamic	0	0	1	0	1	

Favourite type of music, by ethnicity	
uketh2	
1. Autochthonous	1 426
2. Pakistani	58 105
3. Indian	32 56
4. Bangladeshi	6 14
Total	97 601

UK EFFNATIS survey (1999)  
[Penn & Lambert 2009]



## A 'data management' contribution

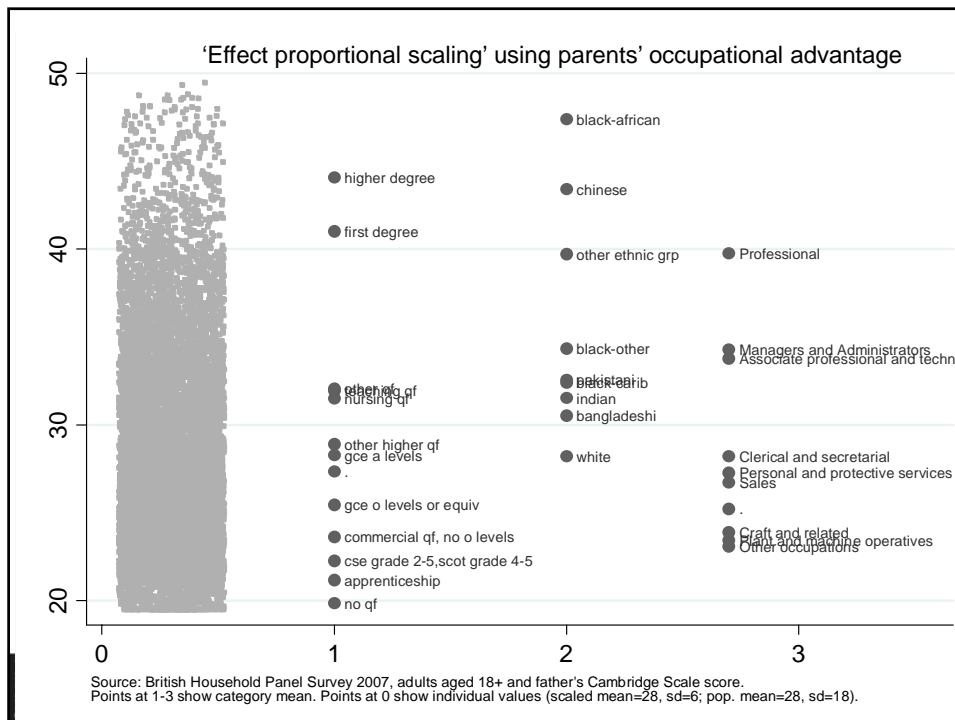
- Preserve information on what was done with categorical data
- Communicate information on what should/could be done

Table 3: Comparison of ethnic group categories in 1991-2001 Census (Simpson and Akinwale, 2006)

Compatible category	1991 categories	2001 categories
White	White	White – all three sub-categories (four in Scotland)
Black Caribbean	Black – Caribbean	Black or Black British – Black Caribbean
Black African	Black – African	Black or Black British – Black African
Indian	Indian	Asian or Asian British – Indian
Pakistani	Pakistani	Asian or Asian British – Pakistani
Bangladeshi	Bangladeshi	Asian or Asian British – Bangladeshi
Chinese	Chinese	Chinese or Other Ethnic Group – Chinese
Other (not comparable over time)	Black – Other Other – Asian Other – Other	Mixed – all four sub-categories (one in Scotland) Black or Black British – Other Black Asian or Asian British – Other Asian Chinese or Other Ethnic Group – Other Ethnic Group

# Standardizing categorical data

- 'Measurement equivalence' (e.g. van Deth, 2003) is often not feasible for complex categorical measures
- For categorical data, equivalence for comparisons is often best approached in terms of meaning equivalence
  - (because of non-linear relations between categories and shifting underlying distributions)
  - (even if measurement equivalence seems possible)
- Arithmetic standardisation offers a convenient form of meaning equivalence by indicating relative position with the structure defined by the current context
- For categorical data, this can be achieved/approximated by **scaling categories** in one or more dimension of difference



## What was that then?

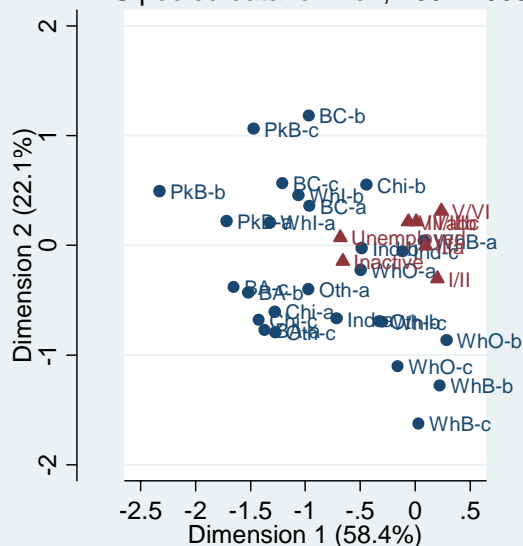
- We can represent categories through positions on a scale
- In turn, we can use position in the dimension as a category score which then plugs into a further analysis (e.g. regression main and interaction effects)

### ..Some options for data on ethnicity..

- Stereotyped Ordered Logistic Regression (SOR) models, summarize dimensions of difference according to regression predictor values [e.g. Lambert and Penn, 2001]
- Geometric data analysis for distances between people, or things [cf. Prandy, 1979; Bennett et al., 2009]
- Assign category scores by hand (*a priori* or by selected average)

## Correspondence analysis dimension scores

LFS pooled data for men, 1991-2005



a = Born in UK; b = Came to UK before 1970; c = came to UK 1970 or later  
N=640295 (Data: Li and Heath, 2008)

## GESDE seeks to promote replicability / transparency with categorical data...

- Document your own recodes / scalings
- Access somebody else's recodes / scalings
- Identify commonly used recodes / scalings  
( & use them..!)

```

Stata Do-File Editor - gemde_bhps_examples.do
File Edit Search Tools
lab_setup.do gemde_bhps_exempl...
capture drop eth2
gen eth2=racel
recode eth2 1/4=1 6/17=2 5 18=3 *=.m /* Recodes to three groups */
replace eth2=3 if (racel==2 & memorig ~= 7) /* Makes 'white irish' a minority only if not in N
capture label drop eth21
label define eth21 1 "White UK" 2 "Black or Asian" 3 "Other white/other"
label values eth2 eth21
numlabel eth21, add

```

## 2) Some old measures – recording & using educational data

		educ4					Total
		-9.00	1.00 Degree	2.00 Diploma	3.00 Higher school or vocational	4.00 School level or below	
Highest educational qualification	-9 Missing or wild	323	0	0	0	0	323
	-7 Proxy respondent	982	0	0	0	0	982
	1 Higher Degree	0	425	0	0	0	425
	2 First Degree	0	1597	0	0	0	1597
	3 Teaching QF	0	0	340	0	0	340
	4 Other Higher QF	0	0	3434	0	0	3434
	5 Nursing QF	0	0	161	0	0	161
	6 GCE A Levels	0	0	0	1811	0	1811
	7 GCE O Levels or Equiv	0	0	0	0	2518	2518
	8 Commercial QF, No O Levels	0	0	0	331	0	331
	9 CSE Grade 2-5, Scot Grade 4-5	0	0	0	0	421	421
	10 Apprenticeship	0	0	0	257	0	257
	11 Other QF	102	0	0	0	0	102
	12 No QF	0	0	0	0	2787	2787
	13 Still At School No QF	138	0	0	0	0	138
Total		1545	2022	3935	2399	5726	15627

## Some educational unit groups

---

### Qualifications listing for the UK:

- British Qualifications (Kogan Page, 2010)
- Qualifications categories of major surveys
  - BHPS, LFS, Census, etc
  - LFS time series standard measure (*used later below*)
- UCAS degree codes

### International schemes

- ISCED: International Standard Classification of Education
  - (cf. Schneider, 2008)
- IPUMS: Census measures over 100 years and 65 countries
- LIS: LFS measures over 50 years and 30 countries

## Educational information resources

---

- In most cases, the original qualifications listing isn't analysed, but is subjected to some recoding transformation
- EIRs capture this data in a recoding matrix or other documentation record

Project specific documentation is often well distributed

e.g. [www.lisproject.org](http://www.lisproject.org)

Address: <http://www.lisproject.org/techdoc/education-level/recoding.htm>

**LIS** Data Access

Standardized Education Recoding

Please read the **methods** and the **instructions** how to use the routine

Successful recoding into variable EDUC

Recoding could not be achieved

Country	Code	Wave I around 1980	Wave II around 1985	Wave III around 1990	Wave IV around 1995
Australia	AU	1981	1985	1989	1995
Austria	AT	.	1987	.	1994/1995/1997
Belgium	BE	.	1985	1988/1992	1995/1997
Brazil	BR	.	.	.	.
Canada	CA	1981	1987	1991	1994/1997
Colombia	BR	.	.	.	.
Czech Republic	CZ	.	.	1992	1996
Denmark	DK	.	1987	1992	1995
Estonia	EE	.	.	.	.
Finland	FI	.	1987	1991	1995
France	FR	1979/1981	1984A/1984B	1989	1994
Germany	DE	1981	1983/1984	1989	1994
Greece	GR	.	.	.	1995
Guatemala	GT	.	.	.	.

E·S·R·C  
ECONOMIC & SOCIAL RESEARCH COUNCIL

However, it's common to find:

- educational categories used which can't be easily replicated
- Users defining their own categories in an *ad hoc* manner
  - Independent approaches (e.g. Lemel, 2002)
  - 'Unequal chances' (Heath and Cheung 2007)
  - Globalife (e.g. Blossfeld et al 2006)

E·S·R·C  
ECONOMIC & SOCIAL RESEARCH COUNCIL

16

## Thinking about 'Old' measures

---

- Good scientific practice to operationalise and use/compare previous measures
  - For most of us, the self-discipline to do this is hard to find!
- GEEDE intends to make this easy and consistent
  - Attract depositors (citation potential)
  - Attract users (easy and better)

## 3) New measures: Issues of functional form

---

- Functional form concerns the representation of main effects, but also, potentially, the use of interaction effects with other controls (e.g. year of birth)

▪ LFS/GHS pooled data (Li and Heath, 2008)

Variable	Obs	Unique	Mean	Min	Max	Label
edcot6esds	2796536	6	2.719703	1	6	education: esds cot: 6cat
edcot2a	2796536	2	.6566631	0	1	Any qualifications level 1 or above
edcot2b	2796536	2	.1658541	0	1	Degree level qualifications
edcot3	2796536	3	1.822517	1	3	Qualifications: None; other; degree
agefted	2513084	27	16.62574	5	31	age left full-time education
ed6age1	2796536	6	16.55295	15.08975	22.15164	Mean age of finishing education
ed6age2	2796348	57	-.0226024	-.6597814	4.66174	Zmean age of finishing education, by level and birth co
year	2796536	34	1991.402	1972	2005	year
yob	2796536	106	1952.176	1884	1989	year
age	2796536	101	39.22577	0	100	age

Regressions with main effects and with interactions

Variable	edcot6	edcot2a	edcot2b	edcot3	agefted	ed6age1	ed6age2
fem	-.511101	-.53133387	-.53729777	-.52834877	-.54463797	-.5232144	-.52405605
year	.05733417	.05950378	.06228331	.05845466	.06085059	.05901138	.06180939
N	596499	596499	596499	596499	596499	596499	596499
ll	-534474.08	-564265.24	-547558.26	-538170.47	-552479.59	-537707.03	-541030.54
r2	.58251235	.53865707	.56378952	.57730597	.55653203	.57796227	.57323304
bic	1069081.1	1128610.3	1095196.3	1076434	1105039	1075493.8	1082140.9

Variable	edcot6_yob	edcot2a_-b	edcot2b_-b	edcot3_yob	agefted_-b	ed6age1_-b	ed6age2_-b
fem	-.51062512	-.53136197	-.53602679	-.52808958	-.54444947	-.52163944	-.52203225
N	596499	596499	596499	596499	596499	596499	596499
ll	-534313.84	-564264.88	-547020.24	-538128.33	-552401.48	-537290.24	-540397.95
r2	.58273659	.53865761	.5645757	.57736569	.55664815	.57855163	.57413725
bic	1068827.2	1128622.9	1094133.6	1076376.3	1104896	1074673.6	1080889

▪ In general terms, scaling of categories is parsimonious, and facilitates including interaction effects

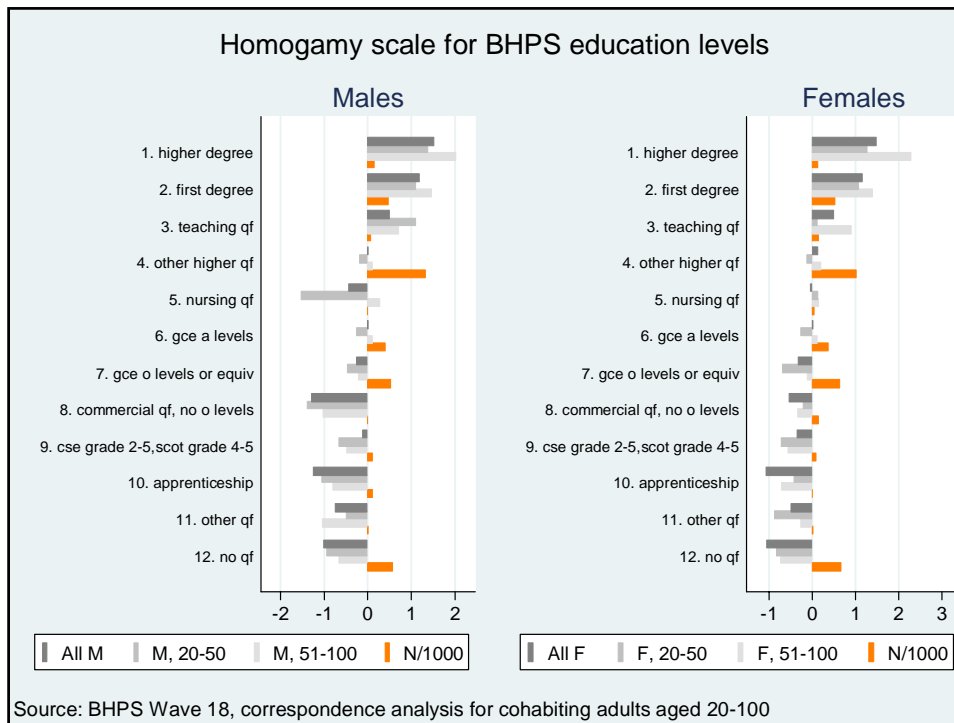
## Different ways to scale...

- Effect proportional scaling (single variables)
- Stereotyped ordered regression scaling (multiple variables)
- Relative value within a context (e.g. birth cohort)
- Linear scaling

		Mean standardised within birth decade*10							
Level	Total	00's	10's	20's	30's	40's	50's	60's	70's
1	15.1	-2	-3	-3	-4	-5	-6	-6	-6
2	16.2	8	6	2	0	-2	-3	-3	-3
3	16.4	2	2	2	2	0	-2	-3	-4
4	16.7	14	8	2	0	-1	-1	0	0
5	19.5	20	21	16	13	12	11	11	13
6	22.2	47	36	31	29	23	20	22	22

## 4) New measures: A CAMSIS model for educational qualifications

- A common way of scaling occupational data is to analyse social interaction patterns between the incumbents of occupations and depict the dimension of social interaction distance as an indicator of stratification
  - CAMSIS approach ([www.camsis.stir.ac.uk](http://www.camsis.stir.ac.uk))
  - Neutral empirical approach, independent of occupational units, comparable across contexts (e.g. Prandy and Jones, 2001)



## Why might this be better than effect proportional scaling?

- Average patterns of social reproduction  
(stable response differences; not conflated with other measures of socio-demographic difference)
- Derive relative scores within contexts  
(gender, birth cohort, region)
- Can work out appropriate scores for unusual/transient units  
*which are more stable than birth cohort specific means*

# Summary

---

- Old measures
  - Finding ways of recording / replicating previous approaches
- New measures
  - Scaling and standardising
  - Attention to parsimony and multivariate context

## References

- Bennett, T., Savage, M., Silva, E. B., Warde, A., Gayo-Cal, M., Wright, D., et al. (2009). *Culture, Class, Distinction*. London: Routledge.
- Blossfeld, H. P., Mills, M., & Bernardi, F. (Eds.). (2006). *Globalization, Uncertainty and Men's Careers: An International Comparison*. Cheltenham: Edward Elgar.
- Heath, A. F., & Cheung, S. Y. (Eds.). (2007). *Unequal Chances: Ethnic Minorities in Western Labour Markets*. Oxford: Oxford University Press.
- Kogan Page Editorial Staff. (2010). *British Qualifications 2010: A Complete Guide to Professional, Vocational and Academic Qualifications in the UK*. London: Kogan Page.
- Lambert, P. S., & Penn, R. D. (2001). *SOR models and Ethnicity data in LIS and LES : Country by Country Report*. Syracuse University, Syracuse, New York 13244-1020: Luxembourg Income Study Paper No. 260, Maxwell School of Citizenship and Public Affairs.
- Lemel, Y., & Noll, H. H. (2002). *Changing Structures of Inequality: A Comparative Perspective*. Montreal: McGill-Queen's University Press.
- Li, Y., & Heath, A. F. (2008). *Socio-Economic Position and Political Support of Black and Ethnic Minority Groups in the United Kingdom, 1972-2005 [computer file]*. 2nd Edition. Colchester, Essex: UK Data Archive [distributor], SN: 5666.
- Penn, R. D., & Lambert, P. S. (2009). *Children of International Migrants in Europe: Comparative Perspectives*. Basingstoke: Palgrave.
- Prandy, K. (1979). Ethnic discrimination in employment and housing. *Ethnic & Racial S*, 2(1), 66-79.
- Prandy, K., & Jones, F. L. (2001). An international comparative analysis of marriage patterns and social stratification. *International Journal of Sociology and Social Policy*, 21, 165-183.
- Schneider, S. L. (2008). *The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries*. Mannheim: MZES.
- Simpson, L., & Akinwale, B. (2006). *Quantifying Stability and Change in Ethnic Group*. Manchester: University of Manchester, CCSR Working Paper 2006-05.
- van Deth, J. W. (2003). Using Published Survey Data. In J. A. Harkness, F. J. R. van de Vijver & P. B. Miller (Eds.), *Cross-Cultural Research Methods* (pp. 222-242). New York: Wiley.