

Comparative scaling of educational categories by homogamy – Analysis of UK data from the BHPS

Paul S. Lambert

University of Stirling

22nd January 2012 [Edition 1.0]

DAMES Node, Technical Paper 2012-1

Technical Papers of the DAMES Node: Data Management through e-Social Science, <http://www.dames.org.uk/publications.html>. DAMES is an ESRC Research Node, Ref: RES-149-25-1066, based at the Universities of Stirling and Glasgow (the National e-Science Centre, www.nesc.ac.uk). The DAMES Node is a component of the ESRC programme in e-Social Science and Digital Social Research (<http://www.digitalsocialresearch.net/>).



Contents:

Introduction	3
Educational categories	4
Social interaction distance analysis	6
Data and methods	7
Results	8
Conclusions	10
References	11

Acknowledgement: The work behind this note was undertaken as part of the ‘GEEDE’ project, ‘Grid Enabled Educational Data Environment’ (www.dames.org.uk/geede), which is itself part of the ESRC funded ‘Data Management through e-Social Science’ research Node (RES-149-25-1066).

1. Introduction

Measures of educational qualifications available on social surveys constitute categorical variables which are often difficult to incorporate into statistical analysis. Problems can arise due to the presence of large numbers of categories of qualifications with an ambiguous status; the existence and advocacy by different analysts of alternative categorical schemes; and institutional reforms of educational systems which lead to correlations between birth cohort and qualification categories. One common solution is to scale categories according to one or more dimensions, where the dimensions are defined by some external criteria. Perhaps the most commonly used scaling is by the average time spent in education for individuals with each level of qualifications, but there are several alternative strategies available (e.g. Buis, 2010).

One interesting alternative means of scaling categories is to use positions within dimensions defined according to the social interaction patterns exhibited by respondents from each category. This note presents a brief illustrative analysis of social distance between educational groups using data from the UK's British Household Panel Survey (BHPS, see University of Essex, 2010). The data files generated, along with a copy of the Stata do file used to conduct this analysis, are posted on the GEEDE website.

2. Educational categories

We used three measures of educational qualifications which can be derived from the BHPS. The first is the original BHPS variable ‘*QFEDHI’ with 12 categories of ‘highest educational qualification’ (see University of Essex, 2010). The other two measures are derived from this and supplementary measures in the BHPS according to a macro deposited on the GEEDE system (see https://dames.cs.stir.ac.uk/liferay/web/geede/ddi/-/ruri/damesZone/home/test4/My%20Curations/test4_DVAOMVXL/test4_DVAOMVXL.ddi). The measures are a derivation of the ISCED scheme as published by Brynin (2003) (with all cases with ‘no qualifications’ imputed to ISCED category 1a/b), and an alternative simplified recoding of the original ‘qfedhi’ measure into four categories (a coding to a low number of categories which reflects a very typical use of educational data in statistical analyses). The tables below show the unweighted number of cases of adult BHPS respondents aged 25 and over with each level of qualification in the 18th wave of the BHPS, after using listwise deletion of cases with missing data on any of the relevant measures. These proportions are indicative, but it is well known that the distribution of cases by qualification levels in a social survey of the adult population such as the BHPS should be expected to vary substantially across other important socio-demographic factors, such as the birth cohort and gender of respondents, as well as due to the possible use of socio-demographically and geographically influenced sampling weights.

‘qfedhi’			
	highest educational qualification	Freq.	Percent
1	higher degree	413	3.7
2	first degree	1,451	12.9
3	teaching qf	298	2.7
4	other higher qf	3,222	28.7
5	nursing qf	135	1.2
6	gce a levels	1,065	9.5
7	gce o levels or equiv	1,619	14.4
8	commercial qf, no o levels	280	2.5
9	cse grade 2-5,scot grade 4-5	265	2.4
10	apprenticeship	206	1.8
11	other qf	64	0.6
12	no qf	2209	19.7
	Total	11227	100.00

'educ4'			
BHPS 4-fold educational level classification			
		Freq.	Percent
	Degree	1,864	16.6
	Diploma	3,655	32.6
	Vocational or higher school level	3,234	28.8
	Low school level or below	2474	22.0
	Total	11227	100.00

'isced_2'			
ISCED 1997, using Brynin (2003) macro, merging 1a and 1b (imputing respondents with 'no qualifications' to category 12)			
		Freq.	Percent
12	1a/1b Incomplete/Elementary	2273	20.2
13	1c Basic vocational	3,761	33.5
21	2a Intermediate vocational (+ intermediate general)	342	3.0
22	2b Intermediate general	1,619	14.4
23	2c General: General maturity certificate	1,069	9.5
24	2d Vocational: Vocational maturity (+with general maturity)	5	0.0
31	3a Lower tertiary	298	2.7
32	3b Higher tertiary	1,866	16.6
	Total	11233	100.00

3. Social interaction distance analysis

The Social Interaction Distance (SID) approach to the analysis of data on social connections can capitalise on established social mechanisms such as those of educational marital homogamy, educational friendship homophily, and intergenerational educational support, in order to discern meaningful structure within the taxonomy of educational qualifications, in terms of social interaction patterns. Common techniques used in this approach include applying correspondence analysis or row-column association models to summarise the cross-tabulation of response categories for two socially connected people (such as pairs connected through marriage, friendship or intergenerationally). Such models generate measures of dimensional structures which help to characterise the patterns of response associated with the various categories (e.g. Wong, 2010). Indeed, the use of these models is already a long-established, and popular contemporary, approach for the scaling of categories of occupations (cf. Prandy and Jones, 2001; Chan, 2010; de Luca et al., 2012; Bessudnov, 2012). Comparable social distance analyses have occasionally been undertaken using educational qualifications in the same style as is shown below (e.g. Duncan-Jones, 1972; Ultee and Luijkx, 1990), but these have been less commonly interpreted as a means of scaling or comparing educational qualifications in themselves.

It is widely acknowledged that educational homogamy and homophily marks all societies and tends in most situations to be somewhat stronger than occupational homogamy. An important part of the structure which defines homogamy and homophily patterns is expected to be the relative social stratification circumstances of individuals: individuals are more likely on average to associate with others of a comparable position in the social stratification structure. Accordingly, when analysing social interaction distance structures, we would expect one dimension, probably the strongest, of the SID structure of educational qualifications, to reflect a hierarchy of social advantage. Accordingly, the scores of each educational qualifications category within that dimension could be expected to make an effective indicator of the relative social advantage typically held by those in possession of each level.

Two important attractions of the SID approach are that (1) the scaling approach offers an opportunity for large numbers of fine-grained categories to be navigated (for instance, the many different categories which might emerge from cross-classifying academic, institutional and vocational measures of educational qualifications); and (2) that the same scaling techniques can be used for several alternative, and apparently non-comparable, categorical indicators of position. For instance, the dimensional scores assigned to several different, more and less detailed, educational taxonomies applied to the same population can be usefully compared in order to understand the relative meaning of each qualification category in the context of any of the taxonomies. SID analysis can thus provide a useful comparative device for summarising and comparing social positions as measured at a detailed level across different social contexts (cf. Lambert et al. 2008).

4. Data and methods

Cohabiting couples in the BHPS dataset were matched in order to generate a record of 9301 distinctive male-female spousal (marriage or cohabitation) educational combinations. Cross-tabulations were performed on that dataset (i.e. husband's education cross-tabulated by wife's education), and modelled through correspondence analysis. The dimensions scores estimated for each educational position were constrained to be equal for the respective rows and columns of the table. Manual inspection of the model results was undertaken to review evidence of other influences upon social connection patterns, and, when relevant, additional selection of the data was applied using the same principles as are ordinarily employed when the same techniques are applied to data on occupations (cf. Prandy and Lambert, 2003; and www.camsis.stir.ac.uk).

The combinations which are analysed are potentially drawn from any year of the BHPS (1991-2008), and the same individuals can potentially contribute more than one spousal combination, if at different points in time they have different highest educational levels, or spouses. However, duplicate combinations (i.e. the same spousal combination with the same qualifications from additional time points) are deleted, and the most recent records in the BHPS period (1991-2008) are prioritised. Details of the compilation of the dataset of couples is available in the attached do file. Prior analysis of social interaction data has suggested that dimensions of the social interaction structure as revealed through association modelling are robust to substantial variations in the sampling of social connections (e.g. Prandy and Lambert, 2003; Chan, 2010). Therefore although the sample used in this analysis cannot be regarded as a random sample of the population of couples in the UK (or of the holders of educational qualifications in the UK), it is nevertheless likely to reflect the same structure of social interaction distance as would have been revealed from a random sample of either population.

5. Results

Three correspondence analyses were conducted for the relation between male and female educational levels within each educational qualifications scheme. Dimension scores were saved and subject to mean standardisation (mean 50, standard deviation 15). The following results were obtained:

	dimension 1	dimension 2	SID score (d1 zscore)
1. higher degree	1.63	1.13	84.90
2. first degree	1.31	0.67	78.09
3. teaching qf	0.64	0.26	63.66
4. other higher qf	0.22	-0.35	54.86
5. nursing qf	0.13	-0.19	52.89
6. gce a levels	0.12	-0.45	52.67
7. gce o levels or equiv	-0.15	-0.46	47.08
8. commercial qf, no o levels	-0.31	-0.41	43.66
9. cse grade 2-5,scot grade 4-5	-0.25	-0.59	44.84
10. apprenticeship	0.00	-1.61	50.18
11. other qf	-0.37	-0.12	42.23
12. no qf	-0.97	0.63	29.56
Dimension 1 %inertia	63.6%		
Dimension 2 % inertia	23.5%		
Total model inertia	0.416		
Notes: Table shows results from correspondence analysis of male-female cohabiting partners' qualifications levels using BHPS microdata from the period 1991-2008. Social Interaction Distance scale scores are a standardisation of the dimension 1 scores. Model shown was a correspondence analysis of all individuals in couples aged over 25, removing all combinations of 'apprenticeship' and 'no qualification' since inspection suggested these formed a 'pseudo diagonal'. Derivation file and tables of data are also available at the GEEDE website.			

	dimension 1	dimension 2	SID score (d1 zscore)
1. Degree	1.53	0.63	82.95
2. Diploma	0.24	-0.43	55.10
3. Vocational or higher school level	-0.18	-0.34	46.23
4. Low school level or below	-0.77	0.57	33.35
Dimension 1 %inertia	81.1%		
Dimension 2 % inertia	17.5%		
Total model inertia	0.291		

Notes: Table shows results from correspondence analysis of male-female cohabiting partners' qualifications levels using BHPS microdata from the period 1991-2008. Social Interaction Distance scale scores are a standardisation of the dimension 1 scores. Model shown was a correspondence analysis of all individuals in couples. Derivation file and tables of data are also available at the GEEDE website.

	dimension 1	dimension 2	SID score (d1 zscore)
11. 1a/1b Incomplete/Elementary	0.88	0.55	30.73
12. 1c Basic vocational	-0.16	-0.45	53.19
21. 2a Intermediate vocational (+ intermediate general)	0.58	0.34	37.41
22. 2b Intermediate general	0.12	-0.48	47.29
23. 2c General: General maturity certificate	-0.15	-0.50	53.01
24. 2d Vocational: Vocational maturity (+with general maturity)	-0.15	-0.50	53.01
31. 3a Lower tertiary	-0.68	0.22	64.47
32. 3b Higher tertiary	-1.43	0.77	80.56
Dimension 1 %inertia	73.7%		
Dimension 2 % inertia	23.0%		
Total model inertia	0.340		

Notes: Table shows results from correspondence analysis of male-female cohabiting partners' qualifications levels using BHPS microdata from the period 1991-2008. Social Interaction Distance scale scores are a standardisation of the dimension 1 scores. Model shown was a correspondence analysis of all individuals in couples aged over 25. Due to sparse numbers of cases categories 23 and 24 were merged in the correspondence analysis. Derivation file and tables of data are also available at the GEEDE website.

6. Conclusions

As is evident from inspection, the SID scale scores shown above appear to reflect a dimension of difference which is characterised by the relative social advantage associated with educational qualifications. As such they could provide a useful means of understanding the circumstances of different types of qualifications, and they can also be feasibly used as scale scores, to be assigned to educational qualifications, in analyses where an indicator of relative educational advantage is required. Indeed, since social interactions are particularly substantively important in defining the social structure, scales based upon their patterns may make particularly good ways of understanding the social structuring of educational qualifications.

Whilst the data used in this example features low numbers of educational categories, social interaction distance analysis may be particularly attractive when large numbers of categories have been measured and an empirical means of ordering them is desirable. The SID approach may also be helpful in circumstances where little other data (on which to base alternative scale scores) is available. In being shown for three different measures, it is also evident that an analysis of social interaction distance offers a useful comparative means of understanding the positioning of categories, in the same relative context, according to various alternative measurement taxonomies (such as comparing the ISCED and 'QFEDHI' taxonomies illustrative above).

In all models shown above, the first dimension is substantially more influential than other dimensions. However the second dimension is also of moderate influence. Speculatively, the second dimension may reveal the propensity of the academic/vocational division to influence patterns of social connections, since in each model it appears related to the division between general academic qualifications and vocational skills. However more work on this pattern would be desirable.

In light of the dimension scores shown, it appears that correspondence analysis proves effective at statistically identifying dimensions of social interaction distance associated with different educational qualifications. Standard errors for the scale scores are not generated by the software package used, but may be appropriate. They could be obtained from conducting the same analysis in software which provides standard errors for association model row and column scores (e.g. the gnm package of Turner and Firth 2007). Various means for assessing the validity of the scale scores generated are available, such as assessing their correlations with other relevant measures. These are not presented here, though it is obvious from the scale values obtained that moderate correlations will be found between the scale scores assigned to educational levels, and other relevant measures.

The derived scores are published as datasets within the GEED portal, see www.dames.org.uk.

References

- Bessudnov, A. (2012). A relational occupational scale for Russia. In P. S. Lambert, R. Connelly, R. M. Blackburn & V. Gayle (Eds.), *Social Stratification: Trends and Processes*. Aldershot: Ashgate.
- Brynin, M. (2003). Using CASMIN: The Effect of Education on Wages in Britain and Germany. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables* (pp. 327-344). New York: Kluwer Academic.
- Buis, M. L. (2010). *Inequality of Educational Outcome and Inequality of Educational Opportunity in the Netherlands during the 20th Century*. Amsterdam: VU University Amsterdam.
- Chan, T. W. (2010). The social status scale: Its construction and properties. In T. W. Chan (Ed.), *Social Status and Cultural Consumption* (pp. 28-56). Cambridge: Cambridge University Press.
- De Luca, D., Meraviglia, C., & Ganzeboom, H. B. G. (2012). Measures and dimensions of occupational stratification: The case of a relational scale for Italy. In P. S. Lambert, Connelly, R., R. M. Blackburn & V. Gayle (Eds.), *Social Stratification: Trends and Processes*. Aldershot: Ashgate.
- Duncan-Jones, P. (1972). Social mobility, canonical scoring and occupational classification. In K. Hope (Ed.), *The Analysis of Social Mobility*. Oxford: Oxford University Press.
- Lambert, P. S., Tan, K. L. L., Gayle, V., Prandy, K., & Bergman, M. M. (2008). The importance of specificity in occupation-based social classifications. *International Journal of Sociology and Social Policy*, 28(5/6), 179-192.
- 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.
- Prandy, K., & Lambert, P. S. (2003). Marriage, Social Distance and the Social Space: An alternative derivation and validation of the Cambridge Scale. *Sociology*, 37(3), 397-411.
- Turner, H., & Firth, D. (2007). *Generalized nonlinear models in R: An overview of the gnm package*. Southampton: ESRC National Centre for Research Methods, Paper 6/07 of the NCRM Working paper series.
- Ultee, W. C., & Luijkx, R. (1990). Educational heterogamy and father-to-son occupational mobility in 23 industrial nations: General societal openness or compensatory strategies of reproduction. *European Sociological Review*, 6, 125-150.
- University of Essex, & Institute for Social and Economic Research. (2010). *British Household Panel Survey: Waves 1-18, 1991-2009 [computer file], 7th Edition*. Colchester, Essex: UK Data Archive [distributor], July 2010, SN: 5151.
- Wong, R. S. K. (2010). *Association Models*. Los Angeles: Sage.