Chapter 40

Educational Mobility, Cohort and Gender; A Latent Class Reanalysis of the Ganzeboom and De Graaf Data

Frank van de Pol

Department of Statistical Methods, CBS, Netherlands

1. Introduction

Mannheim (1928,1929) stressed the idea that generations are formed by common circumstances in the youth of its members (Becker, 1991). One of the aspects that may be considered as typical for a generation is the educational level of its members (Becker, 1987; Becker et al., 1989), as well as their father's educational level. Furthermore, current values that direct educational mobility, and societal structures that channel educational mobility, are reflected in the degree of educational mobility of the members of a generation, conditional on father's educational level. If these features, father's educational level and child's educational mobility, are good indicators for generations, we should be able to attribute cohorts to generations in an unambiguous way. Cohorts should be fully part of a generation or not at all.

In a 1989 paper in „Sociale Wetenschappen“ Ganzeboom and De Graaf analyzed intergenerational educational mobility in the Netherlands for 14 five-year birth-cohorts, from 1891-1895 to 1956-1960. They found that the association between father's education and child's education has decreased. In the present paper we use the same data to look for generational differences, i.e., clusters of cohorts, which display about the same mobility pattern. Differences between sons and daughters will be taken into account.

2. Method and data

To look into this matter we will employ latent class analysis as a sort of cluster analysis. Ganzeboom and De Graaf (1989) brought together data on 22787 subjects from 14 surveys, covering 14 five-year birth cohorts. They recoded the educational level into four categories, 1) primary, 2) junior vocational, 3) general and senior vocational, and 4) vocational colleges and university. For most respondents educational level is the final educational level, for some of the youngest cohorts, however, for a few people educational level may have increased from level 3 to level 4.

The data displayed in Table 1 are summed over the 14 birth cohorts considered. We will analyze 28 of these tables, 14 for men (sons) and 14 for women (daughters). The complete $28 \times 4 \times 4$ table has 448 cells, that is 447 degrees of freedom when sample size is considered fixed.

father's educational level			*child's transition probability,* $\tau_{j\|i} \times 100$			
		$\delta_{i\times 100}$	1. primary	2. junior vocational	3. general, senior voc.	4. voc.col., university
	abs.	%	%			
1. primary	6833	30.0	90.6	6.4	2.3	.6
2. junior vocational	7481	32.8	71.1	20.1	6.4	2.4
3. general, senior vocational	5656	24.8	50.8	23.7	18.6	6.9
4. vocational colleges, university	2817	12.4	31.1	24.5	20.3	24.1

Table 1: Father's educational level, and child's probability of transition to the same or another educational level. Source: Ganzeboom and De Graaf (1989)

A latent class model is used to find out whether a small number, S, of latent tables can describe the H = 28 manifest tables for a specific cohort-gender subpopulation. The model assumes that the table of a specific subpopulation h is a weighted sum of these S latent tables, with weight $\pi_{s|h}$ for table s. Another way to view the same assumption is that table s is valid for subpopulation h with weight $\pi_{s|h}$. When a simple structure would be found, that is when each subpopulation has a high weight $\pi_{s|h}$ on only one of the latent tables, this would support the assertion that generations are homogeneous with respect to educational mobility. To develop notation for these tables, let index i denote the four levels of father's education, mentioned above, and index j denote the four levels of child's education. The proportion in cell {i.j} for manifest table h (h = 1,…H) is written $P_{ij|h}$ ($\Sigma_{ij} P_{ij|h} = 1$). Latent table s (s = 1,…S) is decomposed into a frequency distribution of father's educational level, with proportions $\delta_{i|s}$, and a matrix of transition probabilities $\tau_{j|is}$, as was done with the manifest data in Table 1. The sum of the S tables, weighted with $\pi_{s|h}$, can be observed, $P_{ij|h} = \Sigma_s \pi_{s|h} \delta_{i|s} \tau_{j|is}$ ($\Sigma_s \pi_{s|h} = 1$, $\Sigma_i \delta_{i|s} = 1$ and $\Sigma_j \tau_{j|is} = 1$). In the sample a mixture $p_{ij|h} = \Sigma_s \hat{\pi}_{s|h} \hat{\delta}_{i|s} \hat{\tau}_{j|is}$ is observed.

Given the sampling method, each manifest table is expected to contain proportion γ_h (h = 1,…H) of the sampling population (Ganzeboom and De Graaf, 1989). Due to mortality short-lived persons from the older cohorts are underrepresented in the sample. In order to generalize to the target population of complete Dutch cohorts we have to assume that the same manifest table $P_{ij|h}$ holds for the part of the population that cannot be observed with the present sampling scheme as for the part that can be observed. The model for the proportion in cell {h,i,j} then is

$$P_{hij} = \gamma_h P_{ij|h} = \gamma_h \sum_s \pi_{s|h} \delta_{i|s} \tau_{j|is}$$

$$\text{with} \quad \sum_{hij} P_{hij} = 1 \text{ and } \sum_h \gamma_h = 1 .$$

A similar model for univariate latent budgets instead of bivariate latent tables was proposed by De Leeuw, Van der Heijden and Verboon (1990). In Clogg (1981) and Clogg and Goodman (1985) also latent class models for parallel groups were discussed. In order to obtain maximum likelihood estimates for the target population a correction for the underrepresentation of the older cohorts should be made. We did not do this, however, to avoid increased unaccuracy of the estimates. The model is a special case of the mixed Markov latent class model (Langeheine and Van de Pol, 1990; Van de Pol and Langeheine, 1990; see section 2.2 of chapter 1). Therefore maximum likelihood estimates for the sampling population could be obtained with the computer program PANMARK (Van de Pol, Langeheine and De Jong, 1991).

The likelihood ratio gives the comparison with a saturated model. At first sight, it seems correct to evaluate model fit by looking up the p-value of the likelihood ratio from the theoretical χ^2-distribution with degrees of freedom: number of independent cells – number of parameters. Sample size is large enough. After a closer inspection of the model estimates it was found, however, that many parameters were estimated to be 0, on the boundary of the parameter space. This sheds doubt on the validity of the usual degrees of freedom formula for use with the theoretical χ^2-distribution. Therefore we used a non-naive bootstrap procedure (Langeheine et al., 1996) to evaluate LR model fit. This procedure, which is incorporated in PANMARK, involves drawing many random samples from mobility tables with model-restricted proportions, estimated on the original sample. Such procedures are also called Monte Carlo simulation or parametric bootstrap (Van der Heijden et al., 1996).

3. Results of explorative analyses

When only two latent tables are assumed, the data are described parsimoniously with 85 parameters, which is few compared to the number of parameters that Ganzeboom and De Graaf used. Ganzeboom and De Graaf fitted log-linear models with a constant term and main effects for each of the 28 tables, i.e., 196 parameters, plus some parameters for the interaction term. Our two-table solution, however, does not fit the data well (Table 2).

BIC is Schwartz' information criterion (Sclove, 1987), which should be low for a well-fitting, parsimonious model. BIC suggests the 4-table solution, although the 3-table solution might be interesting too. According to the LR at least four latent tables are necessary for a good model fit. Moreover, interpretation of the 4-table solution is not problematic. Hence we choose the model with four latent tables.

number	*unrestricted*				*restricted*					
of latent tables	para-meters	LR	df	BIC/10	restrictions	para-meters	LR	df	bootstr. p*	BIC/10
2	85	1333	362	24035	2	83	1338	364	.00	24033
3	128	725	319	24017	15	113	730	334	.00	24002
4	171	316	276	24019	48	123	360	324	.47	23975
5	214	216	233	24057	112**	102	1740	345	.00	24092

* At most 500 samples were drawn, unless the standard deviation of the p-estimate became <.05 (as with the 4-table solution: 130 samples).

** to be described in section 4,"testing part of generation theory"

Table 2: Model fit for 2 - 5 latent tables, without and with restrictions on the weights $\pi_{s|h}$

Cohort weights $\pi_{s|h}$ should not fluctuate too much in time; i.e., consecutive cohorts should not load „high-low-high" or „zero-zero-something-zero-zero". This did occur, however, in the unrestricted 4-table solution, especially with the older cohorts, where sample size is small. Therefore equality restrictions were applied in order to eliminate strong fluctuations. (See appendix.) Moreover, equality restrictions were applied to the latent tables, which can also be seen in Table 3. Some restrictions are needed for model identification (De Leeuw, Van der Heijden and Verboon, 1990), others were introduced for smoothing purposes.

latent Table 1: „little mobility“						*latent Table 2: „limited upward mobility“*				
education father		transition prob. child, $\tau_{j\|i1} \times 100$				education f.	transition prob. child, $\tau_{j\|i2} \times 100$			
	$\delta_{i\|1} \times 100$	1.	2.	3.	4.	$\delta_{i\|2} \times 100$	1.	2.	3.	4.
1. primary	93 (1)	97 (3)	2 (2)	0 (1)	1 (0)	83 (3)	29 (3)	45 (2)	20 (2)	7 (1)
2. junior vocational	2 (1)	97 (3)	2 (2)	0 (1)	1 (0)	10 (2)	9 (3)	49 (6)	25 (5)	17 (4)
3. general educat., senior vocationl	3 (1)	22 (6)	18 (7)	51 (8)	8 (9)	5 (1)	11 (4)	20 (6)	40 (7)	30 (7)
4. voc. colleges university	3 (1)	22 (6)	18 (7)	51 (8)	8 (9)	3 (1)	0 (-)	0 (-)	0 (-)	100 (-)

latent Table 3: „fewer with prim. educ. only“						*latent Table 4: „meritocracy“*				
education father		transition prob. child, $\tau_{j\|i3} \times 100$				education f.	transition prob. child, $\tau_{j\|i4} \times 100$			
	$\delta_{i\|3} \times 100$	1.	2.	3.	4.	$\delta_{i\|4} \times 100$	1.	2.	3.	4.
1. primary	55 (5)	5 (12)	79 (10)	16 (3)	0 (-)	33 (3)	3 (1)	10 (2)	64 (2)	23 (2)
2. junior vocational	24 (3)	17 (2)	56 (3)	25 (3)	2 (2)	35 (1)	3 (1)	24 (2)	45 (2)	28 (2)
3. general educat., senior vocationl	13 (2)	8 (2)	40 (4)	41 (4)	11 (3)	21 (1)	3 (1)	10 (2)	52 (2)	35 (2)
4. voc. colleges university	8 (1)	1 (2)	28 (4)	38 (5)	33 (5)	11 (1)	3 (1)	10 (2)	30 (3)	57 (4)

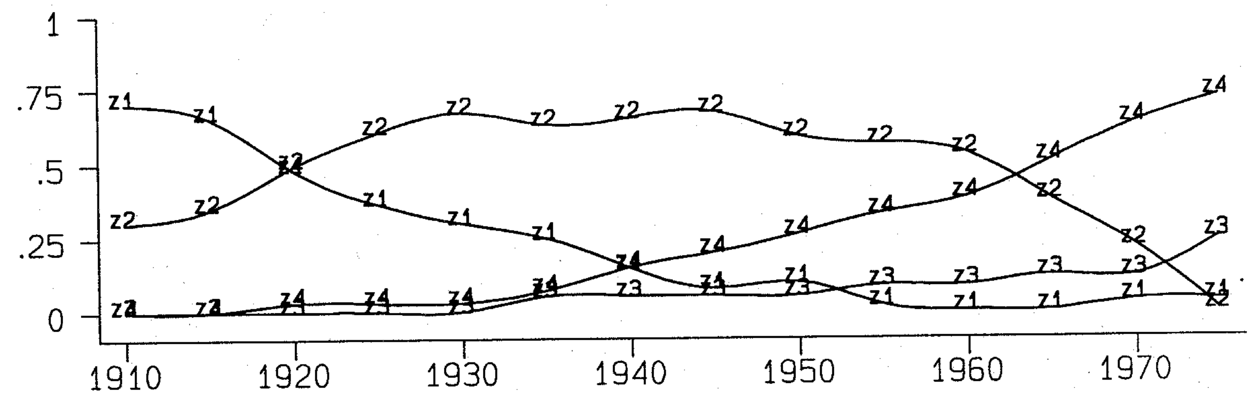
Table 3: Latent tables of restricted 4-table solution (with standard errors).

The fit of the restricted 4-table solution is good, according to the bootstrap-procedure. If the model is true, about 47% of the samples would fit less well than the original sample does, which means that the model has a good fit. With the theoretical χ^2-distribution we would have estimated this percentage to be only 8%. Standard errors of parameters should be interpreted carefully, as a coarse indication of accuracy.

Table 3 shows the four latent tables, the one which was most prevalent for the older cohorts first, the most modern one last. Obviously, the level of father's education has increased throughout the years. Moreover, the latent tables also show the finding of Ganzeboom and De Graaf (1989) that the association of father's education and child's education has decreased in time. To find out whether educational mobility can demarcate generations we should inspect the matrix of weights $\pi_{s|h}$ of cohorts, which are available both for men and women. These weights are plotted in Figures 1 and 2.

The year of reference of each cohort is shifted 17 years because this age is thought to be relatively important for the final educational level attained. Cohort weights were connected with a „spline" using STATA (1988).

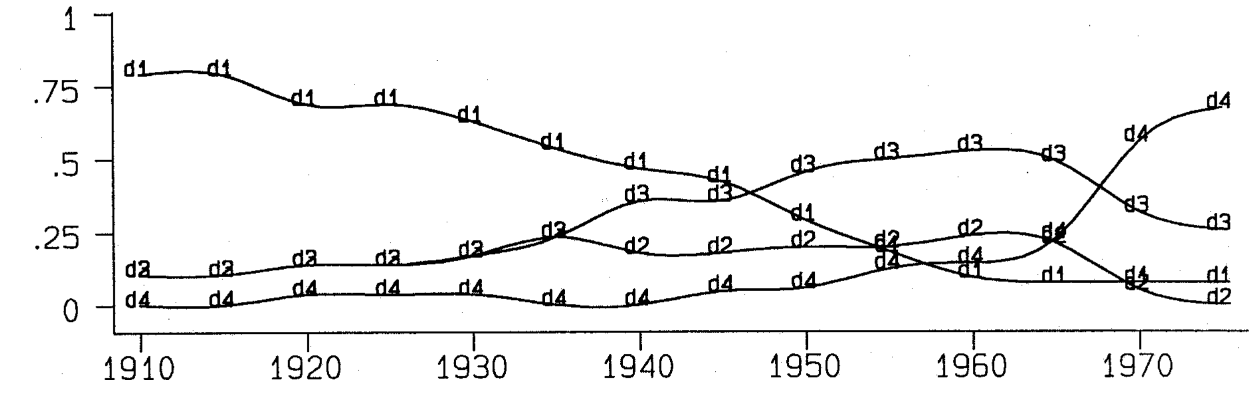
Figures 1 and 2 do not show „natural" demarcation lines between generations. For some cohorts one latent table is the most important, for other cohorts two or three latent tables are about equally important.



year when son became 17 years of age

z1: latent Table 1, „little mobility" z2: latent Table 2, „limited upward mobility"
z3: latent Table 3, „fewer with primary education only" z4: latent Table 4, „meritocracy"

Figure 1: Weights $\pi_{s|h}$ in restricted 4-table solution for men (sons)



year when daughter became 17 years of age

d1: latent Table 1, „little mobility" d2: latent Table 2, „limited upward mobility"
d3: latent Table 3, „fewer with primary education only" d4: latent Table 4, „meritocracy"

Figure 2: Weights $\pi_{s|h}$ in restricted 4-table solution for women (daughters)

For men latent Table 2, which may be labeled as „limited upward mobility", has been the dominant pattern during 40 years. Before this time Table 1 was prevalent, with low educational level and little mobility, and after the „limited upward mobility" period Table 4 is dominant, with high mobility and a high educational level („meritocracy“). Latent Table 3 is specific for women.

For women latent Table 1, low educational level and little mobility, has been the dominant pattern for the first half of the period covered, much longer than for men. Then latent Table 3 takes over, while latent Table 2 is valid for a minority. Table 3, which is specific for women, differs from Table 2 by fewer fathers with low educational level, and by a higher probability to obtain junior vocational education. The Netherlands educational system has a junior vocational school where women learn to become a housewive. This school used to be popular some decades ago. Finally, later than for men, also for women latent Table 4, „meritocracy“, becomes the most important table for educational mobility.

4. Testing part of generation theory

Our conclusion on the basis of an exploratory analysis is that no sharp demarcation lines between generations can be found with four latent educational mobility tables. However, rather than estimating cohort weights, one could fix them also in advance. For this we use Becker et al.'s (1989) classification into four generations: 1) birth cohorts 1910-1930 (17 years of age around 1937), 2) 1930-1940 (17 around 1952), 3) 1940-1955 (17 around 1965), and 4) 1955-1970. (The 1956-1960 cohort in the Ganzeboom and De Graaf data was 17 in about 1975.) We added the 1891-1910 cohorts as a fifth generation. For each of these five generations one latent table is assumed ($s = 1,\ldots,5$). Cohort weights $\pi_{s|h}$ are 1 if cohort h belongs to generation s, otherwise 0.

This „generation theory model“ has 345 degrees of freedom (102 parameters). However, the model does not fit the data, LR = 1740 with bootstrap p = .00, mostly because gender differences are not taken into account. In the past century, males seem to have adopted a new type of educational mobility sooner than females. Also in terms of parsimony the generation theory model is not preferable, BIC = 240923 (Table 1). When confronted with gender-specific data, the current pattern of educational mobility does not seem to be very important for the formation of a gender-inspecific generation.

It is clear that generation theory needs some refinement with respect to its validity for educational mobility. Which changes are in order is a matter that we would rather leave to generation theory specialists. Figures 1 and 2 show at any rate that, for each gender, cohorts can be classified into clusters by calling those clusters after the dominant latent mobility table.

Acknowledgements

The views expressed in this paper are those of the author and do not necessarily reflect the policies of the Statistics Netherlands. The author thanks B. Bakker, H. Becker, P. de Graaf and J. Winkels for ideas and opinions. A previous version of this paper appeared in: Dynamics of cohort and generations research, H. Becker, ed., 1992, pp. 369-379 (Thesis publ., Utrecht, The Netherlands).

Appendix: Parameter restrictions and estimates of table weights; restricted 4-table solution

	restrictions*				estimates (with standard errors)							
	latent table, *s*				latent table, *s*							
birth cohort, *h*	1	2	3	4	1		2		3		4	
men (sons)												
1891-1895 (+17)	0	0	1	1	.70	(.07)	.30	(.07)	.00	(fixed)	.00	(fixed)
1896-1900 (+17)	0	0	1	1	.65	(.06)	.35	(.06)	.00	(fixed)	.00	(fixed)
1901-1905 (+17)	0	0	-1	2	.48	(.05)	.49	(.06)	.00	(fixed)	.03	(.04)
1906-1910 (+17)	0	0	-1	2	.36	(.04)	.61	(.05)	.00	(fixed)	.03	(.04)
1911-1915 (+17)	0	0	-1	2	.30	(.04)	.67	(.05)	.00	(fixed)	.03	(.04)
1916-1920 (+17)	0	0	3	0	.25	(.03)	.63	(.05)	.05	(.03)	.07	(.04)
1921-1925 (+17)	0	0	3	0	.14	(.03)	.65	(.05)	.05	(.03)	.15	(.05)
1926-1930 (+17)	0	0	3	0	.08	(.03)	.67	(.05)	.05	(.03)	.20	(.05)
1931-1935 (+17)	0	0	3	0	.10	(.02)	.59	(.04)	.05	(.03)	.26	(.04)
1936-1940 (+17)	0	0	4	0	.02	(.02)	.57	(.05)	.09	(.04)	.33	(.04)
1941-1945 (+17)	0	0	4	0	.00	(bound)	.53	(.05)	.09	(.04)	.38	(.04)
1946-1950 (+17)	0	0	0	0	.00	(bound)	.38	(.04)	.12	(.05)	.51	(.03)
1951-1955 (+17)	5	0	0	0	.03	(.02)	.22	(.06)	.12	(.05)	.63	(.03)
1956-1960 (+17)	5	-1	0	0	.03	(.02)	.00	(fixed)	.24	(.04)	.72	(.04)
women (daughters)												
1891-1895 (+17)	0	6	6	-1	.79	(.05)	.10	(.02)	.10	(.02)	.00	(fixed)
1896-1900 (+17)	0	6	6	-1	.79	(.05)	.10	(.02)	.10	(.02)	.00	(fixed)
1901-1905 (+17)	0	7	7	8	.69	(.04)	.14	(.02)	.14	(.02)	.04	(.02)
1906-1910 (+17)	0	7	7	8	.69	(.04)	.14	(.02)	.14	(.02)	.04	(.02)
1911-1915 (+17)	0	9	9	8	.63	(.04)	.17	(.02)	.17	(.02)	.04	(.02)
1916-1920 (+17)	0	10	10	-1	.54	(.04)	.23	(.02)	.23	(.02)	.00	(fixed)
1921-1925 (+17)	0	12	13	-1	.47	(.03)	.18	(.04)	.36	(.04)	.00	(fixed)
1926-1930 (+17)	0	12	13	0	.42	(.03)	.18	(.04)	.36	(.04)	.05	(.05)
1931-1935 (+17)	0	14	0	0	.29	(.04)	.20	(.06)	.46	(.06)	.06	(.03)
1936-1940 (+17)	0	14	0	0	.18	(.04)	.20	(.06)	.50	(.06)	.12	(.03)
1941-1945 (+17)	0	0	0	0	.09	(.04)	.24	(.07)	.53	(.06)	.14	(.03)
1946-1950 (+17)	15	0	0	0	.07	(.03)	.21	(.06)	.49	(.06)	.22	(.03)
1951-1955 (+17)	15	0	0	0	.07	(.03)	.05	(.05)	.32	(.05)	.56	(.04)
1956-1960 (+17)	15	-1	0	0	.07	(.03)	.00	(fixed)	.25	(.04)	.67	(.04)
weighted average					.20		.34		.21		.24	

*0: free, -1: fixed at zero, 2: $\pi_{4|3} = \pi_{4|4} = \pi_{4|5}$ 3: $\pi_{3|6} = \pi_{3|7} = \pi_{3|8} = \pi_{3|9}$ 4: etc.

Table A.1: Latent table weights, $\pi_{s|h}$

father's educational level, i	latent Table 1 child's educational level, j 1	2	3	4	latent Table 2 child's educational level, j 1	2	3	4
1. primary	21	22	23	24	0	0	0	0
2. junior vocational	21	22	23	24	0	0	0	0
3. general, senior vocat.	25	26	27	28	0	0	0	0
4. voc.colleges, university	25	26	27	28	31	31	31	0

father's educational level, i	latent Table 3 child's educational level, j 1	2	3	4	latent Table 4 child's educational level, j 1	2	3	4
1. primary	0	0	0	0	51	52	0	0
2. junior vocational	0	0	0	0	51	0	0	0
3. general, senior vocat	0	0	0	0	51	52	0	0
4. voc.colleges, university	0	0	0	0	51	52	0	0

0: free, 21: $\tau_{1|11} = \tau_{1|21}$, 22: $\tau_{2|11} = \tau_{2|21}$ 51: $\tau_{1|14} = \tau_{1|24} = \tau_{1|34} = \tau_{1|44}$, etc.

Table A.2: Restrictions on transition probabilities, from father's education i to child's education j, $\tau_{j|is}$

References

Becker, H.A. (1978). Generaties en sociale dynamiek. In: W.E.A. van Beek, (ed*.), Sociologisch en antropologisch jaarboek 1987* (Van Loghum Slaterus, Deventer), pp. 112-129.

Becker, H.A. (1991). Aspecten van generaties en cohorten: een overzicht van het onderzoeksprogramma. *Sociologische Gids*, *38*, pp. 212-226.

Becker, H.A., Boerman, F.A. and Hermkens, P.L.J. (1989). Trendbreuken en hun demografische neerslag. In: H.A. Becker, and P.L.J. Hermkens, (eds.), *Oude naast nieuwe generaties. Maatschappelijke positie en waardenoriëntaties van ouderen vergeleken met jongeren* (ISOR, RU Utrecht), pp. 101-126.

Clogg, C.C. (1981). Latent structure models of mobility. *American Journal of Sociology, 86*, pp. 836-868.

Clogg, C.C., and Goodman, L.A. (1985). Simultaneous latent structure analysis in several groups. In: N.B. Tuma (ed.), *Sociological Methodology 1985* (San Francisco, Jossey-Bass), pp. 81-110.

De Leeuw, J., van der Heijden, P.G.M., and Verboon, P. (1990). A latent time budget model. *Statistica Neerlandica*, *44*, pp. 1-22.

Ganzeboom, H.B.G., and de Graaf, P.M. (1989). Intergenerationele opleidingsmobiliteit in Nederland van geboortecohorten 1891-1960. *Sociale Wetenschappen*, *32*, pp. 263-278.

Langeheine, R., and van de Pol, F. (1990). A unifying framework for Markov modeling in discrete space and discrete time. *Sociological Methods & Research*, *18*, pp. 416-441.

Langeheine, R., Pannekoek, J. and van de Pol, F. (1996). Bootstrapping goodness-of-fit measures in categorical data analysis. *To be published in: Sociological Methods & Research, 24*.

Mannheim (1928, 1929). Das Problem der Generationen. *Kölner Vierteljahreshefte für Soziologie, 7*, pp. 157-189, pp. 309-330.

Sclove, S.L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*, pp. 333-343.

STATA (1988). *Reference manual release 2* (Computing Resource Center, Los Angeles).

Van de Pol, F., and Langeheine, R. (1990). Mixed Markov latent class models. In: C.C. Clogg, (ed.), *Sociological Methodology 1990* (Blackwell, Oxford), pp. 213-247.

Van de Pol, F., Langeheine, R., and de Jong, W. (1991*). PANMARK 3 user's manual; PANel analysis using MARKov chains, a latent class analysis program.* (Voorburg; e-mail: fpol@cbs.nl).

Van der Heijden, P.,'t Hart, H., and Dessens, J. (1996*) A parametric bootstrap procedure to perform tests in a LCA of anti-social behaviour*. In: this volume.