On the Shapes of Social Change

James A. Davis
NORC and University of Chicago
September, 2009

GSS Social Change Report No. 56

Introduction

We are (as usual) in an unusual period of social change and modern replication surveys are designed to monitor it. Granted we know a bit about whether means are increasing or decreasing, little is known about *shape* – whether the trajectories are straight line, wavy, monotonic with plateaus, or what? Major students of racial attitudes (Schuman et. al., 1997, Schwartz, 1967) fit trends nicely with straight lines, without bumps for historic events or economic cycles. But social change is not limited to racial attitudes. Hence the purpose of this essay is to assess the linearity and non-linearity of important social trends.

Linearity in regressions is usually assumed, occasionally inspected, sometimes manufactured, and rarely interpreted. This essay suggests one would be rewarded to go beyond arbitrary decisions about "linear v. non-linear" to a scrutiny of "degree of linearity". . Why?

1. The inference problem is clear. If the X→Y function is nonlinear, the true "curve" must cross the best fitting line at least once[1]. At the crossings the error variance will be zero. The residuals will then rise in both directions as X moves away from the crossing point, invalidating the assumption of homoscedascicity. Although very similar to the dependent dummy problem, this has received little attention from practical methodologists. Nevertheless, I won't discuss it because (a) I don't have any simple suggestions for dealing with it and (b) Modern probability replication samples have so many cases that statistical inference is essentially a ritual rather than a useful interrogation.

2. Common sense suggests real world relationships are hardly ever *perfectly* linear. If so, and if we can fit a plausible non-linear function, we can improve our *R* squares routinely.

3. Most important, perhaps, is that non-linearities are interesting. If an effect of educational attainment is linear; all we actually learn is "the more the more". If, however, it bends sharply after, say, eight years of schooling we suspect there is some input from primary education that is different from that of later

[1] I think the actual number is twice, but I don't have a proof.

levels. If we find that cohorts born around 1950 (who reached late adolescence in the 1960s) have distinctive attitudes, we find support for the media's bloviating about "generation this and generation that".

Methods

Regression I. (Xc→Ymean)^2

Textbook cures for non-linearity boil down to curve fitting or transformations (logs, powers, roots. etc). Each produces a shape which hopefully matches the Y means at each value of X. Easier said than done and the resulting functions usually defy substantive interpretation.

A much simpler procedure:[2]

1. Divide X, the interval level independent variable, into as many equal spaced categories as possible, without compromising category reliability. Call it Xc[3]
2. Find the *mean* Y (dependent variable) each level of Xc
3. Regress the Y means on Xc. Call the result (Xc→Ymean.)
4. Square the value.[4] ..

(Xc→Ymean)^2 is a measure of linearity. When it is very large, the relationship is highly linear. The argument is this: in the regression, the program attempts to fit a least squares straight line to the sequence of category means. If the X→Y function is actually perfectly linear, each mean will be on that line and the correlation will be +/- 1.00. To the extent the Y means stray from the line, the function is non-linear and the correlation is less than 1.00. We have switched linearity from a platonic quality to a variable. Interpretation of

[2] All this can be done with the push of a button using the "aggregate" command in SPSS or its equivalent in other statistical packages. The procedure is also known as "effect-proportional scaling " (Treiman, p. 257-58)

[3] One hesitates adding neologisms to notation but we will have multiple versions of X→ Y, and they are clumsy to distinguish verbally

.

[4] Squarins is not mathematically necessary but it (1) eliminates coefficient signs and (2) spreads out the distribution of high magnitudes.

its values will be treated later. For now, here are some examples from the 1972-2006 General Social Survey (The survey is described below).

(Figure 1 here - Graphs are all in very preliminary form.)

Figure 1 has four panels. The first displays four examples where **r**=.96, $\mathbf{r^2}$ =.98, the Second four examples where **r**= .75, $\mathbf{r^2}$ = .56, the third **r** = .55, $\mathbf{r^2}$ = .30, and the last **r** = .02**,** $\mathbf{r^2}$ = .00. The horizontal variables are Age, Cohort (year of birth0, Education, or year. The vertical values are for selected variables in this analysis (See Appendix II.)
The story is clear: At .96/.98 all four lines are essentially straight, at .76/87 all four bend but have an unambiguous direction, at .56/.75 U-shapes turn up, at .30/.55 three of the four have U-shapes, and at .00/.02 we see one U-shape and three shapeless wiggles.

Regression II. (Xdum→Y)

Regressions with dummy variable predictors provide a second tool for assessing linearity. One proceeds as follows:

1) Create a set of dummy variables, Cdum, comprising the C category levels of X.[5]
2) Regress Y on the dummies, dropping one as usual
3) The resulting **r** is (Xdum→Y)
4) Correlate (XdumY) and (X→Y)

The *predicted* values produced by (Xdum→Y) may be construed as a fittable "curve" analogous to a straight line or parabola or whatever. . The "function" will defy mathematical description but, although nameless, it fits X→Y like the proverbial glove, better than any

[5] Variables such as age or income in dollars must be collapsed to produce practical dummies. In theory information is lost. My experience, however, has been that the effect on coefficients is trivial provided one has more than a handful of dummies.

possible alternative. That is, the difference between the "observed" and "predicted" Y *means* (not cases) will all be zero – since the prediction *is* the category's Y mean.

(Xdum→Y) thus has an important property:

> *It produces the largest bivariate correlation between X and Y of any possible function.*

Running saved values of (Xdum→Y) against Y gives the second measure of linearity, logically similar to $(Xc \rightarrow Ymean)^2$. If the relationship is perfectly linear, the correlation will be plus or minus 1.0. (the dummy predictions match the linear predictions.) As values stray from the line, the correlations decline and if there is no directionality at all, the *r* will be zero[6].

I prefer $(XcYmean)^2$ because it only requires on calculations but (XdumY) will be shown to be quite useful.

Inspection

Neither regression tells us anything about the actual shape when the linearity coefficients are small. Low values can mean chaotic jiggling, U-shapes, step functions, s-curves, and so on. To see what is going on one must examine plots as in Figure 1. [7]

Coding shapes is not easy, especially when the line contains "*ears*". This is not a technical term but the concept emerged from the inspections reported here. In several cases the graphs appeared to be a reasonable line or curve with exceptions at either extreme. Observe YEAR and GRASS in Table 1 $\mathbf{r^2} = .30$. Beginning around 1978 we see a routine inverted U, but before that the line moves up. In other words, attitudes toward marijuana

[6] Whether to square either coefficient is a matter of taste.
[7] Since SPSS automatically adjusts the plot so the vertical scale ends slightly above and slightly below the extreme, plotted values its plots are deceptive when judging *magnitudes*. (XdumY), however, does this nicely.

.basically became more favorable in the 1980s and less favorable in the 90's BUT the years prior to 1980 do not fit this model. The distinction between "ears" and step functions is murky but I chose the former when the discrepant line was not horizontal. Ears, of course, can occur at either the right or left hand side of the series.

I ended up with the following types:

Linear: essentially straight (e.g. $\mathbf{r^2}$ = .96 in Table 1)
Bowed: curved, no bends, no plateaus ($\mathbf{r^2}$ =.75, Year and Abdefect in Figure 1)
Step/Plateau: a group of essentially identical values followed or preceded by a Linear or bowed sequence. (e.g. Figure 4)
U or inverted U: (e.g. $\mathbf{r^2}$ = .56, Dmarried and Cohort in Figure 1)
Complex (?) Nil or pattern-less (e.g. $\mathbf{r^2}$ = .00, Age and Happy in Figure 1)

Data

I chose some 136 variables from the NORC General Social Survey (GSS)[8] Appendix 2 gives descriptions of each.. Four items – the key predictors of social change – are treated as independent and the remainder regressed on them[9]. The independent variables are Year, Age, Cohort, (Birth Year) and Education.

Year

Between 1972 and 2006 NORC fielded 26 versions of the GSS for a total of 51,020 cases. The GSS was planned as an annual study but vicissitudes of funding made this impossible. Between 1972 and 1993 studies were carried out every year save for 1979, 1982, and 1992. This first series had an average yearly N of 1,547. Beginning in 1994 the project shifted to a biennial design with a doubled sample size (mean N = 3090).

YEAR in the cumulative GSS file may be treated as a continuous series, but possible complications arise from gaps among dependent items. In theory the GSS consists of a "core" of permanent items plus "one-shot" supplements on various topics. In practice, not every core item appears every year for two reasons: (1) to make precious space many core items were placed in a rotating plan such that prior to 1988 they appeared at two year intervals with one year gaps. (2) The project occasionally added new items of sudden interest and/or removed items that seemed outmoded.

---

[8] The GSS is a once-annual, now-biennial, area probability design, personal interview sample of U.S., English speaking (a handful of Spanish only speakers were added in 2006 but are excluded here) householders ages 18 and older. Completion rates range a bit below 75 percent. For the analyses here the data have been weighted to make them representative of adult individuals not households. The National Science Foundation has provided continual partial support.

[9] The analyses are oblivious to causal order. Noter VdumY is *not* perfectly symmetrical. For example, (AgeDum→Educc) gives an **r** of .270. Running the opposite (EduccDum→Age) yields an **r** of .300.

. **Age**

Age is divided into 26 categories (18-19 to 89+) etc. for comparisons with Year. The grouped measure correlates +.985 with the raw values of age.

Cohort (Birth Year)

Cohort (year of birth ranging from 1883 to 1998) was also divided into 26 equal frequency categories for comparability with Year and Age. Strictly speaking the intervals are not perfectly equal since they were created to make essentially equal Ns in each category not equal distances. . Nevertheless, the **r** between raw and grouped versions is +.984.

Education

The GSS measures education as "years completed"[10] from 0 to 20 (Mnemonic = EDUC). Because cases are thin at the lowest levels – especially in later years - I grouped 0 through 5 as 2.5 and 6 through 7 as 6.5. In addition I combined 19 and 20 as 19.5 since "year" is ambiguous in many graduate programs. . This gives a total of fourteen levels with the mnemonic Educc (EDUC Collapsed).

Dependent Items

Of the 132 dependent variables, 53 (40%) appeared in all 26 years, 95 (72%) in 20 or more years, 124 (94%) in 15-19 years and all at least 11 years. Since linearity could be sensitive to the number of time points selected, I ran the $(YEARcYdum)^2$ coefficients against: first year, last year, span= first year minus last year, N, and number of data points. None showed a large, consistent or reliable relationship (N = 133)[11].

[10] The survey provides a second measure, "DEGREE", or highest degree. DEGREE and "EDUC" correlate +.849. I chose EDUC because it gives a finer breakdown.

[11] Appendix II lists 136 items. Four are the predictors, 132 the dependent items. In the major runs, however, three of the four predictors are treated as dependent. (See note 9) Unless otherwise noted, the analysis are

The dependent items were chosen to cover a variety of topics and a large range of years. The roughly grouped topics are: family attitudes, family structure, free speech, gender roles, geography, life/death, parental family, politics, racial attitudes among whites, religion, sex behavior, sex norms, sociability, socio-economic status, values, and well- being. Appendix 2 lists the specific items[12].

Results

Linearity Distributions

Table 1 displays the univariate distributions of linearity coefficients $(VcYmean)^2$ for Age, Period, Cohort, and Education.

From the viewpoint of perfect linearity, the values are not high. Ten percent or less are "perfectly straight" (.95+) and the medians are mostly between .50 and .60 (half the year to year variance in means is linear).

based on 135 coefficients.

[12] Among the sociologically salient topics under-represented might be, networks, national political issues, media and internet usage, and cultural consumption.

Table 1.
Cumulative Distributions (Percentaged) of Linearity Coefficients (r^2)

Cumulative %	Age	Year	Cohort	Education
95+	9	4	5	10
85-94	23	10	22	44
75-84	38	27	41	61
65-74	46	38	50	66
55-64	51	52	56	72
0-54	100	100	100	100
Median	.564	.562	.651	.819
Mean	.532	.477	.551	.677
N	135	135	135	135

Before drawing a firm conclusion it is necessary to consider strength because stronger relationships are straighter: the bivariate **r**'s between VdumYmean^2 (linearity) and VdumY (strength) are Age= .422 Period = .475, Cohort = .418, Education = .600 N=133. This is presumably because weak relationships have larger error variances which generate random departures from linearity. Table 2 summarizes the distributions for VdumY, our measure of strength.

Table 2
Bivariate Distributions [r] for (VdumY)

	Age	Year	Cohort	Education
0.75	.217	.140	.266	.151
Median	.162	.108	.181	.180
Mean	.144	.090	.145	.151
0.25	.052	.059	.067	.091
N	135	135	135	135

Given the large sample sizes, (raw Ns per item range from 15,111 to 51,020) all but a handful of coefficients are reliable but the magnitudes are not impressive. . The Age, Cohort, and Education relationships are typically (mean. median) close to .15, while Year correlations are about a third smaller.

To see linearity among the *stronger* associations, Table 3 displays the four linearity $(VdumYmean)^2$ distributions for the 25 items with the largest values of (VdumY).

Table 3
Cumulative Bivariate Distributions of $(VdumYmean)^2$
(25 Strongest Relationships)

Cumulative %	Age	Year	Cohort	Education
90+	36	48	24	88
70-89	60	76	60	92
50-69	72	88	88	96
<50	100	100	100	100
N	25	25	25	25

Table 3 suggests:

`

While perfectly linear patterns are rare, save for Education, a modest majority Are sufficiently straight as to justify standard OLS.

A small minority (4 to 12 percent) are clearly non-linear.

Education relationships are the most linear for Cohort relationships least.

These are the key findings of the report

To the extent these results are representative of the stronger trend relationships, in the majority of cases the OLS assumption of linearity is harmless, (although non-linear approaches would boost R^2) but in a small, but non-trivial, percentage the linearity assumption would distort the true relationship

Patterns

As useful as $(VdumYmean)^2$ may be, the coefficient itself does not tell us anything specific about the shapes. In the respectable minority of cases with low linearities inspection of shape is necessary Tables 4a – 4d display the author- coded shapes for the 25 strongest relationships of Age, Period, Cohort, and Education, using the following symbols

OK = straight, linear
B = bowed
ST = step
U = u-shaped
? = complex or nil

As noted above, several of the distributions contain two or more points at the extreme differing from the overall pattern. Lacking a technical term, I will call them "ears" and designate them by "L" or "R" for left and right. Thus L B means a bowed curve with exceptions at the left (lowest level of X).

Table 4a.
Shapes of Distributions in Table 3: AGE

Mnemonic	$(Vdum\ Ymean)^2$	(VdumY)	OK	B0W	STEP	U	?
COHORT	.986	.829	OK				
XMOVIE	.984	.272	L OK				
PORNLAW	.982	.272	OK				
PARED	.974	.370	L OK				
UNDEMP	.972	.319	L OK				
Dpamdif	.960	.412	OK				
PREMARSX	.949	.288	L OK				
Liberal	.947	.276	L OK R				
SOCFREND	.931	.339	OK				
SOCBAR	.893	.348	L OK				
RACMAR	.884	.277		B			
Sumath	.867	.290		B			
Sumall	.841	.297		B			

FEHELP	.817	.330	B			
Summil	.814	.259	L B			
Mar1 (Single)	.626	.586	B			
Mar5 (Widowed)	.551	.457	B			
Educc	.520	.269			U	
Dwork	.473	.493			U	
EARNRS	.472	.448				?
Athome	.419	.445			L U	
Dwifwork	.391	.363			U	
SEXFREQ	.297	.458			U	
Dmarried	.193	.370			U	
REALINC	.011	.257			L U	

Table 4b.
Shapes of Distributions in Table 3: YEAR

Mnemonic	(Vdum Ymean)^2	(VdumY)	OK	B0W	STEP	U	?
COHORT	.960	.511	OK				
Devdivorcd	.960	.156	OK				
Dfamdif	.960	.156	OK				
RACPUSH	.960	.321	OK				
RACSEG	.960	.228	OK				
Educc	.912	.218	OK				
FEHELP	.912	.257	OK				
FEHOME	.912	.192	OK				
Pared	.912	.227	OK				
RACMAR	.912	.214	OK				
Sumall	.912	.165	OK				
Sumhomo	.912	.218	OK				
BUSING	.846	.177	OK				
RACDIF2	.846	.159	OK				
RACOPEN	.846	.253		B			
Sumcom	.846	.155	OK				
Athome	.757	.156	OK				
Dwifwork	.757	.219		B			
FEPOL	.757	.221			ST		
COURTS	.672	.163				U	

HOMOSEX	.672	.190		ST		
SUICIDE1	.672	.170		ST		
Liberal	.423	.270			U	
Nats4	.123	.207				?
CAPPUN	.023	.186		ST		

Table 4c.
Shapes of Distributions in Table 3: COHORT

Mnemonic	(Vdum Ymean)^2	(VdumY)	OK	B0W	STEP	U	?
Dfamdif	.976	.420	OK				
Pared	.976	.414	OK				
RACOPEN	.964	.436	OK				
AGE	.951	.828	OK				
RACSEG	.949	.478	OK				
Liberal	.918	.411		B			
FEHELP	.852	.363		B			
YEAR	.839	.455					?
MAR1 (single)	.808	.481		B			
FEHOME	.805	.338		B			
Sumcom	.789	.370			ST		
Summil	.787	.340			ST		
Sumhomo	.776	.321		B			
Dwifwork	.774	.332				U	
Sumath	.738	.386			ST		
RACPRES	.682	.342		B			
Dwork	.679	.457			ST		
Mar5 (widowed)	.645	.373		B			
Educc	.645	.332			ST		
EARNRS	.615	.319			ST R		
SEXSEX1	.549	.564	OK				
POLVIEWS	.517	.531		B			
WORDSUM	.094	.325			U		
REALINC	.054	.375			U		
PRESTG80	.010	.316			U		

Table 4d.
Shapes of Distributions in Table 3: EDUCATION

Mnemonic	(Vdum Ymean)^2	(VdumY)	OK	B0W	STEP	U	?
WORDSUM	.986	.482	OK				
JOBMEANS	.982	.306	OK				
Liberal	.956	.357	L OK				
Sumcom	.956	.404	L OK				
FINRELA	.953	.299	OK				
Sumall	.951	.458	L OK				
RACSEG	.947	.326	L OK				
Sumath	.945	.417	L OK				
Sumhomo	.943	.393	L OK				
Incmpc	.941	.316	OK				
REALINC	.941	.372	OK				
Summil	.941	.344	L OK				
Pared	.933	.495		B			
Sibs	.925	.338		B			
RACPUSH	.924	.378	L OK	B			
CLASS	.918	.316	OK				
FEWORK	.914	.292					
FEHOME	.912	.380		L B			
RACMAR	.912	.392		L B			
Papres16	.904	.353	L OK				
PRESTIGE	.903	.624		B			
Dpovline	.901	.343		B			
PRESTG80	.863	.553	L OK				
COHORT	.643	.368				U	
AGE	.446	.300				U	

Table 5 collects the patterns in Table 4a – 4d.

TABLE 5.
Summary of Patterns in Table 4

	Age	Year	Cohort	Education
Linear (OK)	10	16	6	13
Bowed (B)	7	3	8	10
Step (ST)	0	3	6	0
U-Shaped (U)	7	0	4	2
Complex/Nil (?)	1	3	1	0
	25	25	25	25
"Ears"				
L	8	0	0	13
R	1	0	1	0

Combining Linear, Bowed, and Step as monotonic and hence r appropriate for linear OLS, the four predictors are essentially similar in terms of monotonicity. About three quarters of their strongest relationships could be described reasonably by a straight line. The non-linearities' shapes however, differ from predictor to predictor.

AGE has seven u-shapes – six of which are the familiar "life cycle" values declining on both sides of the middle years. They are all "objective" variables: labor force status, number of children in household, working wife if married, sex frequency, currently married and family income. Figure 2 shows the life cycle in late 20th century America – the average of the z scores for each of the six. It starts with -.01 at age 19, rises to +.30 at age 42 and then declines steadily to - .93 at age 83. While Age→Education has a u-shape, formal schooling seldom continues past age 25. This will come up again when we consider Cohort.

(Figure 2 here)

In addition to the familiar life cycle pattern, there is a different shape for more "subjective" items. Age has ten "ears", nine at the left (lower) end. SOCBAR is typical. Socializing at bars increases from 18 to 24 and then declines steadily with age[13]. Taken together the ears suggest the life cycle in Figure 2 is not the whole story. Before the mid twenties one sees some quite different age patterns among the youngest adults. Figure 3 plots the age trajectory for five late adolescence items (Liberal, PREMARSX, SOCBAR, Summil, and XMOVIE.). The line is the mean of the five items normalized.

(Figure 3 here)

Figure 3 supports the common assumption that social attitudes are far from fixed by age sixteen.

Turning to YEAR perhaps the most interesting feature is the absence of "humps" that might suggest multi-item "periods" such as the "Clinton era". The nine non-linear items in Table 4b have different turning points and maxima: FEPOL may have hit a ceiling (see below), COURTS has a U-shape with "too harsh" increasing until 1978 and decreasing after 1994, HOMOSEX shows a large increase in tolerance after 1989, and the rest (Liberal, GRASS, Nats4, and CAPPUN) display patterns not easily classified.

Remembering the Year correlations are relatively lower, the conclusion is: in contrast with Age, Cohort, and Education, Year relationships are "weaker and straighter".

[13] One might construe this as a u-shape except that the maximum for the left branch is much lower than the other.

. COHORT (birth year) is notable for its many non-linear but monotonic shapes (14 of 20 in Table 5). One immediately thinks of the endlessly touted, but seldom documented unique attitudes of the "baby boomers" versus the allegedly less liberal attitudes of their predecessors and successors (Davis 2002, 2004). The actual patterns are a bit different.

Before considering these attitudes we note a sociologically important non-linearity in schooling. Beginning with the birth cohort of 1948, mean years of educational attainment ceased their long- term increase. The finding has received considerable attention. Goldin and Katz (2008), for example, argue the plateau has had a strong impact on American inequality. (For an alternative view on changing inequality, see Bartels, 2008). The U shape for WORDSUM (vocabulary score) is consistent with this view although I'd be inclined to view the U shapes for REALINC and PRESTIG80 as heavily life cycle driven. Figure 4 plots the cohort means for Education and WORDSUM for respondents 25 and older[14].

(Figure 4 here)

Almost all attitude items show some sort of slope change toward the end of the GSS era but the patterns vary.

The race and gender role items generally show a bow pattern with a de*cellerating* liberal increase. Close inspection leads to caution. Almost all of the race [15] and gender[16] items are dangerously near their highest possible scores in the later cohorts. This suggests ceiling effect artifacts. Whether progress has slowed down or the GSS items can't capture change at the highly liberal end is unknown (The GSS was designed in the early 1970s using items all of which had appeared in *earlier* national surveys.) At the least, one might say the

[14] Respondents younger than 25 may still be completing their educations
[15] RACMAR, RACOPEN, RACSEG, RACPRES, RACDIF2, RACPUSH, Tipping point
[16] FECHLD, FEHELP, FEHOME, FEPO, FEPRES, FRPRESCH, FEWORK

birth cohorts around 1950 saw the final evaporation of self-admitted crude racism and sexism[17].

For three clusters of "liberal/conservative" items, marginals are moderate enough to allow close scrutiny: sex norms, abortion, and free expression. Table 6 displays the results.

Table 6
Cohort and Shape for Sex Norms, Abortion, and Free Expression Items

Item	Cluster	Shape	Cohort of Inflexion Point
PORNLAW	Sex	Step	1947
ABDEFECT	Abortion	U	1947
Sumhomo	Free Expression	Bow	1947
Sumath	Free Expression	Step	1947-1952
PREMARSX	Sex	Step	1950 (at ceiling?)
HOMOSEX	Sex	Bow*	1952
Sumcom	Free Expression	Step	1952
ABSINGLE	Abortion	U	1952
Summil	Free Expression	Step	1952
XMARSEX	Sex	U	1952
ABNOMORE	Abortion	U	1953
ABPOOR	Abortion	U	1953
ABHLTH	Abortion	U	1958
ABRAPE	Abortion	U	1963
SEXEDUC	Sex	Bow	none (at ceiling?)
TEENSEX	Sex	OK	none

***** The mean increases up to 1952, drops and then increases

All items show increasing "liberalism" up to an inflection point, after which the trend is "boom" era (1945-1960), so the later the birth, the more liberal the response. Subsequently thirteen of fifteen either hit a plateau (at below presumable ceiling levels) or reverse direction toward lesser liberalism.

[17] Both generalizations, while socially encouraging, are sociologically challenging. Once crude racism and sexism are off the table, it is most unclear exactly what designers of racial attitude trend questionnaires should be asking.

Figure 5 graphs three rough scales against Cohort.[18]

What to make of the pattern is not obvious. Since those born in the 1950's reached late adolescence in the 1960's it is tempting to invoke the social turmoil of the 1960's. (Figure 3 is consistent with the hypothesis that key social attitudes are fixed in the late teens) If so, the effect should be a *temporary* bump (Davis 2004), i.e. a U shape. Among the attitude items in Table 6 only the abortion series has a U shape and abortion was not a prominent 1960's controversy. The stalling of education (Figure 3) suggests an alternative explanation. Since liberalism generally increases with schooling, stalled education might lead to a stalling in the liberalism trends of items in table 6.[19] Table 7 tests this idea.

[18] All items were normalized and then averaged. Abortion = ABDEFECR, ABHLTH, ABNOMORE, ABSINGLE, ABRAPE, Civil Liberties = Sumath, Sumcom, Sumhomo, Summil, Sumrace. Sex = HOMOSEX, PORNLAW, PREMARSX, SEXEDUC, TEENSEX, XMARSEX. All items were coded so + = "libearak".

[19] A *possible* explanation for the abortion exception: - to a much greater extent than free expression or progressive sex norms, abortion has elicited highly organized opposition in the last few decades.

Table 7.
Effects of Earlier and Later Cohorts on Three Attitude Scales

COHORT -> SCALE (betas)

Cohort = 1883 - 1951			Cohort = 1952 - 1981	
		Abortion Scale		Difference
Bivariate	.0290		-.0649	
Net of Education	-.0132		-.0673	-.0541
Difference	-.0422		-.0024	
N = 23.012			9,261	
		Civil Liberties Scale		
Bivariate	.1775		.0052	
Net of Education	.1093		.0028	-.1065
Difference	-.0682		-.0024	
N = 19,593			23,012	
		Sex Norms Scale		
Bivariate	.1092		.0450	
Net of Education	.1050		.0436	-.0614
Difference	-.0042		-.0014	
N= 27,774			13,501	

Age 25 and older only, 1883-1951 coefficients are divided by 1.9464 to compensate for cohort difference in standard deviations.

The "differences" (e.g. -.0132 - .0673) tell the story, In all three cases the impact of Education is smaller in the post 1951 cohorts – though the effect is trivial for Sex Norms In other words, the gain in liberalism is less in the later cohorts, net of education – so the drop

can not be explained by education. Education plateau makes a discernable contribution to the two of the three liberalism plateaus, but the part it plays is small.

The fourth predictor, EDUCATION, (Table 5) has the fewest bends and twists as 23 of 25 relationships are monotonic. However, inspection of the plots reveals 10 "ears", all on the left. In each case the line is horizontal prior to nine years. Apparently elementary education has less impact on these items than secondary or tertiary. Figures 6 and 7 illustrate these mild non-linearities for five attitudes[20] and five SES measures.[21]

Methodological Implications?

A practical approach which guarantees better predictions (larger values of **r)** merits discussion beyond social change research. As a start, Table 8 shows the gains for the 540 bivariate analyses treated above.

Table 8.
Linear Versus Dummy Variable Bivariates

	Age	Year	Cohort	Education
	(a) Raw Regression			
75%	.188	.105	.220	.228
Median	.091	.057	.099	.131
25%	.023	.029	.031	.045
	(b) VarDumY minus VarY absolute			
75%	.040	.036	.077	.037
Median	.022	.021	.030	.020
25%	.012	.012	.010	.011
	(c) VarDumY divided by VarY absolute			
75%	2.1	2.3	2.6	1.7
Median	1.3	1.4	1.4	1.2

[20] FEHOME, Liberal, RACMAR, RACPUSH, RACSEG
[21] CLASS, PRESTG80, PRESTIGE, REALINC, Incmpc

25%	1.1	1.1	1.1	1.1
N	133	133	133	133

Panel (a) shows a typical OLS bivariate to have an **r** between .06 and 13, panel (b) says that the dummy variable approach adds 2 to 3 correlation points; panel (c) says that the dummy variable approach improves correlations from 20 to 40 percent (small absolute values produce impressive ratios).

Should one therefore always shift to dummies? Maybe, but maybe not. First, as noted above, larger **r squares** have higher priority in methods classes than in published research (after all, 025 squared = .000625). Second, as in studies of economic fluctuations, the story is often in the linear gain amidst the convolutions. Even the u-shaped life cycle scale correlates -279 with Age; we all end up lower than when we began. Third, there are costs – each additional calculation gives an opportunity for typing errors and requires a careful explanation in the text.

The author's opinion: We should already be following the textbook advice to examine bivariate plots. U-shapes should be tested with VdumY and transformed unless the strength is non-trivial. Scattered bows and plateaus should be left as is unless the research question focuses on the size of $\mathbf{R^{2.}}$. If, however, several items have a meaningful shapes (e.g. Cohort before and after 1947) they should be transformed to dummies and the results discussed in the report.

Conclusion

Social changes in mass phenomena seem to be neither as melodramatic in shape as pop sociology (periods, cycles, "the XXX generation") would suggest are or as slim and

straight as routine research assumes. More often than not linearity analysis is harmless. However, non -linearities are common enough and substantively interesting enough to merit careful scrutiny. To this author, the key question is not heteroscedasisity but whether the analyst is telling the correct story about what is going on.

****
*

REFENCES CITED

Bartels, Larry M. 2008. *Unequal Democracy.* Princeton NJ. Princeton University. Press

Goldin, Claudia and Lawrence F. Katz. 2008. *The Race Between Education and Technology.* Cambridge MA. Harvard

Davis, James A. 1992. "Changeable Weather in a Cooling Climate Atop the Liberal Plateau." *Public Opinion Quarterly. 56: 261-306*

_____ 2004. "Did Growing Up in the 1960's Leave a permanent Mark on Attitudes and Values?" *Public Opinion Quarterly.* 68: 161-183

Schuman, Howard, Charlotte Steeh, Lawrence Bobo and Maria Krysan . (1997) *Racial Attitudes in America: Revised Edition.* Cambridge MA. Harvard.

Schwartz, Mildred A. 1967. *Trends in White Attitudes Toward Negroes.* Chicago. NORC..

Treiman, Donald J. (2009) *Quantitative Data Analysis. Jossey-Bass/Wiley.*

****
*

APPENDIX I. NOTATION

V = Variable
X= Independent variable
Y= Dependent Variable, raw values

Xc = X collapsed into c categories
Xdum = X recoded as a set of dummy variables
Ymean = mean of Y for a category of X

X→Y = OLS regression, raw data
Xc→Ymean = correlation between levels of X and their means on Y
Xdum→Y = correlation between X dummies and raw data Y

$(XdumYmean)^2$ = linearity measure

.

APPENDIX II. Variables

Left hand column = variable name
CAPS = GSS mnemonic, Lower Case = recode
X = Blacks excluded
Content = Paraphrase of item topic
Age, Period, Cohort, Education: results of analyses in main text
Blank = VdumY < .224 (..224 sq. = .050)
Regular = $(VdumYmean)^2$ if VdumY >= .224 & < .316 (..316sq.. = .100)
<u>Underline</u> = $(VdumYmean)^2$ if VdumY >= .316 & < .447 (.447 sq = .200)
Bold = $(VdumYmean)^2$ if VdumY >= .447

Mnemonic	Content	Age	Period	Cohort	Education
1. ABDEFECT	Allow abortion: Fetus is defective				
2. ABHLTH	Allow abortion: Mother's health in danger				
3. ABNOMORE	Allow abortion: Doesn't want more				
4. ABPOOR	Allow abortion: Family is poor				
5. ABRAPE	Allow abortion: She was raped				
6. ABSINGLE	Allow abortion: Single, prefers no marriage				
7. ADULTS	Persons 18+ in household				
8. Age14	AGE in 5 year intervals	Inap		**.951**	.446
9. AGED	Should elders live with adult children			.943	
10. AGEWED	Age at first marriage			.533	
11. Athome	Persons <18 in household (BABIES+PRETEENS+TEENS)	<u>.419</u>		.420	
12. ATTEND	Frequency of Church Attendance				
13. Belt1 (SRCBELT)	Resides in center, largest metros				

14. Belt2 (SRCBELT)	Resides in center, medium metros				
15. Belt3 (SRCBELT)	Resides in suburb of Belt1				
16. Belt4 (SRCBELT)	Resides in suburb of Belt2				
17. Belt5 (SRCBELT)	Resides in small town				
18. Belt6 (SRCBELT)	Resides in rural area				
19. BIBLE	Is bible inerrant				.964
20. BORN	Born in US				
21. BUSING-X	Busing children for racial integration			.729	
22. CAPPUN	Death penalty for murderers				
23. CHILDS	Children even born				
24. CHLDIDEL	Ideal number of children				
25. CLASS	Self-rated social standing				.918
26. COHORT	Year of birth	**.986**	**.960**	Inap	.643
27. COMMUN	Attitudes towards communism as a system				
28. COURTS	Harshness of local courts				
29. Dblack	Dummy for RACE				
30. Dchristian	Dummy for RELIG = Prot & Catholic				
31. Divdivorcd	Ever divorced if ever married				
32. Dfamdif	If not with parents at age 16	.960	.960	.976	
	Parents both dead v. parents divorced				
33. Dindep PARTID	Neither Democrat nor Republic Independent				
34. DIVLAW	Divorce should be easier or harder				
35. DMARRIED	(MARITAL)				
36. Dmidwest	(REGION) Current residence	.193		.511	
37. Dmidwest16	(REG16) Residence at 16				
38. Dneast	(REGION) Current residence				
39. Dneast16	(REG16) Residence at 16				
	NewEngland & MidAtlantic	.972			
40. Dpovline	(POVLINE) Above/below Federal poverty line				.901
41. Dsouth	(REGION) Current residence South				
42. Dsouth16	(REG16) Residence at 16				
	S.Atlantic & S.Central				
43. Dtax	(TAX) Federal Income Tax				
44. Dunemp	Unemployed in the past 10 years				
45. Dwest	(REGION) Current residence				
46. Dwest16	(REG16) Residence at 16				
	Mountain & Pacific				
47. Dwifwork	In labor force or if not married female	.391		.774	.943
48. Dwork	(WRKSTAT) In labor force	**.473**		.679	.876
49. EARNRS	Number employed in household	**.472**		.615	
50. EDUCc	(EDUC) Years of schooling	.520		.645	Inap
	collapsed (0-5=3, 6-7=6.5)				
51. EQWLTH	Should government equalize incomes				

52. ETHNUM	Strength of ethnic identification				
53. Farm16	(REG16) Living on a farm at age 16	.916		.901	.799
54. FEAR	Fearful place nearby				
55. FECHILD	Working mom doesn't hurt children				
56. FEHELP	Should put husband's career first	.817	.912	.852	.941
57. FEHOME	Women should stay home	.797		.805	.912
58. FEPOL	Women not suited for politics			.839	
59. FEPRES	Vote for woman presidential candidate			.792	
60. FEPRESCH	Preschooler suffers if mother works			.884	
61. FEWORK	Should married women work			.659	.914
62. FINALTER*	Own finance worse/same/better				
63. FINRELA	Self-rated income				.953
64. FUND16	Fundamentalism of R's church, age 16				
65. GRASS	Legalize marijuana				
66. GUNLAW	Require gun permits				
67. HAPMAR	Happiness of own marriage				
68. HAPPY	Self-rated happiness				
69. HEALTH	Self-rated health				
70. Helpblkres	(HELPBLK) See end of table				
71. Helpsum	(HELPNOT, HELPPOOR, HELPSICK) Welfare state index. See end of table				
72. HOMOSEX	Homosexuality always wrong			.863	.937
73. HOMPOP	Total persons in household				
74. Incmpc	(REALINC) Household income per capita <18's = 1/2 person				.941
75. JOBFIND	Easy/hard to find a job				
76. JOBINC	High income a job priority				
77. JOBLOSE	Likely to lose current job				
78. JOBMEANS	Meaningful work a job priority				
79. LETDIE1	Euthenasia for incurable patients				.982
80. Liberal	Grab bag index of liberal opinions (33) See end of table	.947	.423	.918	.956
81. LIFE					.953
82. Mar1	MARITAL single never married	**.626**		**.808**	
83. Mar2	MARITAL/DIVORCE married, never divorced	.159			
84. Mar3	MARITAL divorced				
85. Mar4	MARITAL/DIVORCE married, been divorced				
86. Mar5	MARITAL widowed	**.551**		.645	
87. MOBILE16	Moves since age 16				
88. Natres-X	(NATBLACK) See end of table				
89. Nats4	(NATEDUC, NATHLTH, NATCITY, NATCITY) Welfare state index. See end of table			.870	
90. OWNGUN	Gun in home				

91. PAPRES16	Father's occupational prestige old scale				.904
92. Paranoia	(mean on FAIR, HELPFUL, TRUST)				.937
	See end of table				**.933**
93. Pared	Mean of parents' years of schooling	.974	.912	.976	
94. Party2	(PARTYID) Independent				
95. POLVIEWS	Self-rated liberal v. conservative			**.517**	
96. PORNLAW	Legalize pornography	.982			
97. PREMARSEX	Premarital sex - how wrong	.949		.824	
98. PRESTG80	Prestige of r's occupation - new scale			.010	**.863**
99. PRESTG	Prestige of r's occupation - old scale				**.903**
100. RACDIF1-X	Race inequality due to - discrimination				
101. RACDIF2-X	Race inequality due to - inborn				
102. RACDIF3-X	Race inequality due to - education			.169	
103. RACDIF4-X	Race inequality due to - willpower				
104. RACMAR-X	Legalize inter-racial marriage	.884			.912
105. RACOPEN-X	Vote on open housing		.846	.964	
106. RACPRES-X	Vote for black presidential candidate			.682	
107. RACPUSH-X	Blacks shouldn't push where not wanted		.960	.982	.924
108. RACSEG-X	Whites have right to segregated neighborhood				
109. REALINC	Family annual income in 1968 dollars	.011		.054	.941
110. RELITEN	Intensity of religiosity				
111. RICHWORK	Work/quit if suddenly rich				
112. SEX	Gender				
113. SEXED	Sex education in public schools				.787
114. SEXFREQ	Frequency of sex	**.297**		.865	
115. Sexsex1	Sexual partners hetero-to-homo			**.549**	
116. Sibsr	Total brothers and sisters				.925
117. SOCBAR	Frequency: evenings at bars	.893			
118. SOCFRIEND	Frequency: evenings with friends	.931		.949	
119. SOCOMMUN	Frequency: evenings with neighbors				.630
120. SOCREL	Frequency: evenings with relatives				
121. SPANKING	OK to spank children				
122. SUICIDE1	Allow suicide - incurable disease				
123. SUICIDE4	Allow suicide - tired of living				
124. SumAll	Index: 15 free speech (Stouffer) items	.841			**.951**
125. SumAth	Index: Free expression for anti-religious	.867		.738	
126. SumCom	Index: Free expression for communist	.870		.789	.956
127. SumHomo	Index: Free expression for homosexual	.776		.776	.943
128. SumMil	Index: Free speech for militarist	.814		.787	
129. SumRac	Index: Free speech for racist			.410	.941
130. TEENSEX	Sex among teens 14-16 - how wrong				
131. ThinkObey*	Priority for a child "Obedience" vs. "Think for his/her self"			.466	.970

132. Tippingpoint					
133. Yearprob	Date of Survey		Inap	**.839**	.834
134. WORSUM	Total correct on vocabulary test			.094	**.986**
135. XMARSEX	How wrong extra-marital sex				
136. XMOVIE	Seen X-rated film this year	.984			

****

*

70. HelpBlkRes HELPBLK (Should government increase aid to blacks) residualized on welfare state help items (HELPNOT, HELPSICK ,HELPPOOR

71 Helpsum Pro Welfare state index – sum on HELPNOT, HELPSICK, HELPPOOR

80. Liberal "Grab bag" index of liberalism items (Nats4 – see below, religious fundamentalism, religious intensity, premarital sex, extramarital sex, Blacks shouldn't push, racial inter-marriage, spending on military (FUND, RELITEN, PREMARSEX, XMARSEX, RACPUSH, RACMAR, NATARMS)

88 NATRES-X Should government spend more on Blacks (NATBLACK) residualized on spending for cities, education, environment, health.

89 NATS4 Federal spending index: for or against spending on cities (NATCITY) education (NATEDUC) environment (NATENVIR), health (NATHEAL)

90. Paranoia Three item index based on "trust" items (FAIR, HELPFUL, TRUST)

91. Tipping point-X: Guttman style scale based on "Would you object" to sending your children to a school with FEW, HALF, MOSTLY Black students.

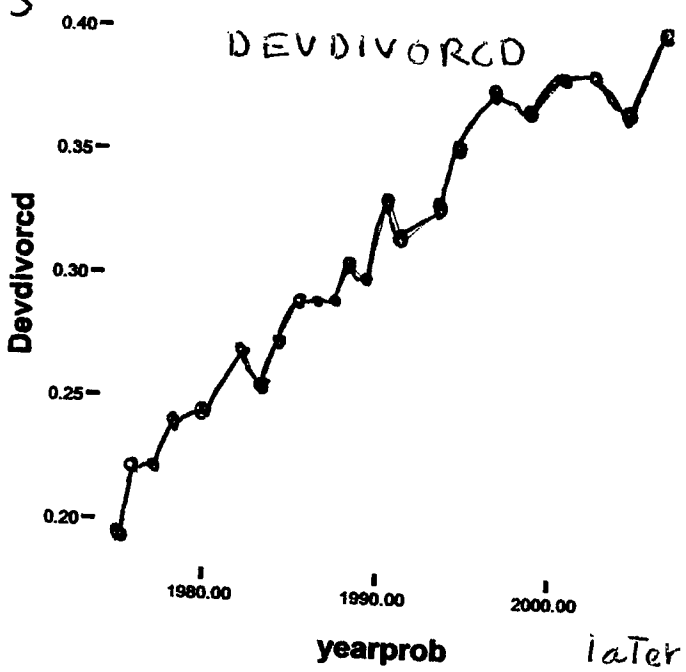
Figures 1- 7

To accompany "On the Shapes of Social Change"

Figure 1 $r^2 = .96$ $r = .98$

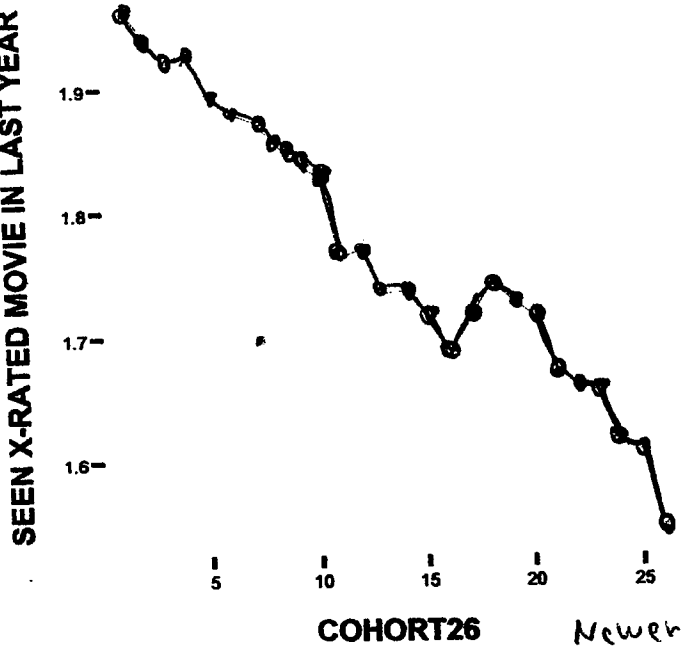
WRONG PREMARSX

YES DEVDIVORCD



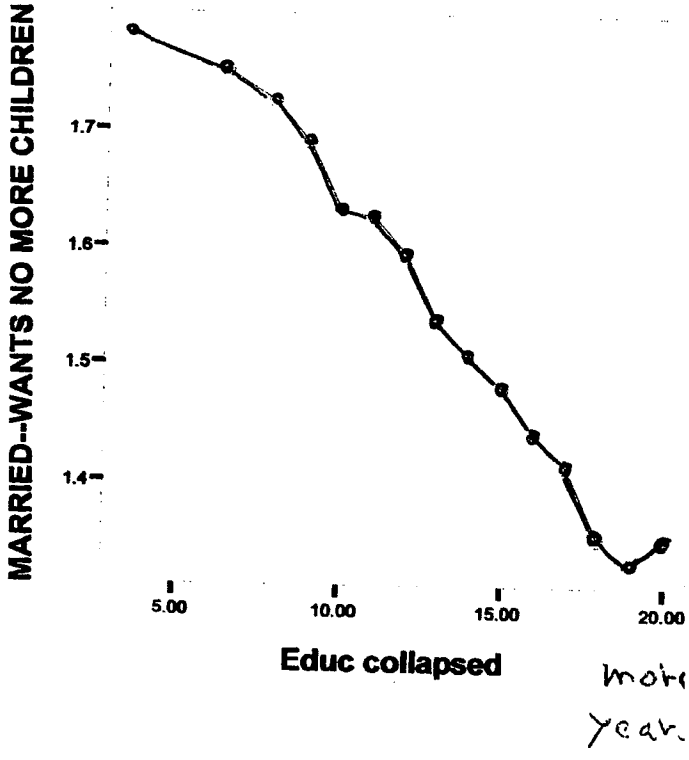
XMOVIE

NO



ABNOMORE

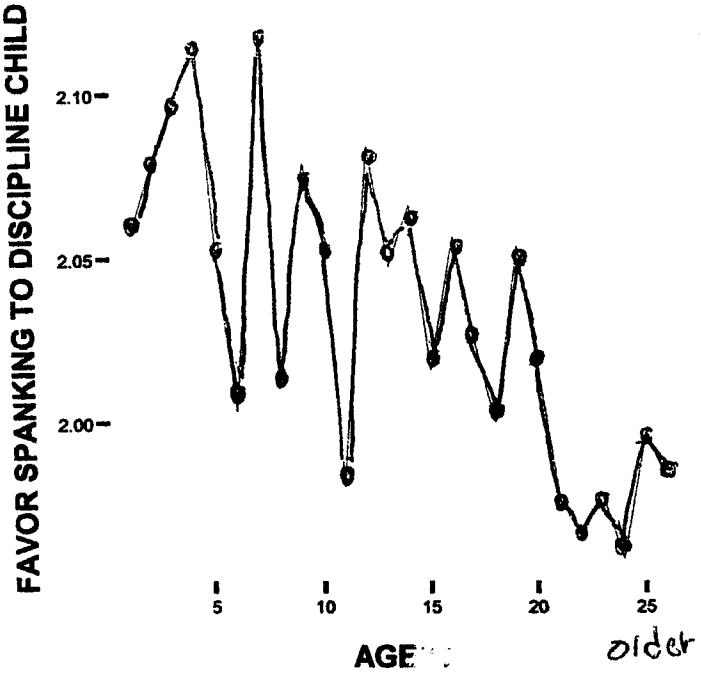
NO



$r^2 = .56$ $r = .75$

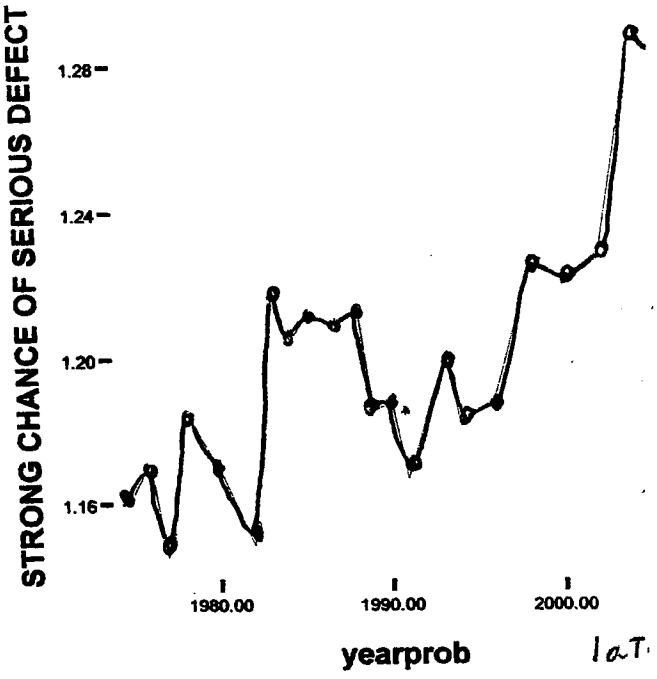
SPANKING

YES



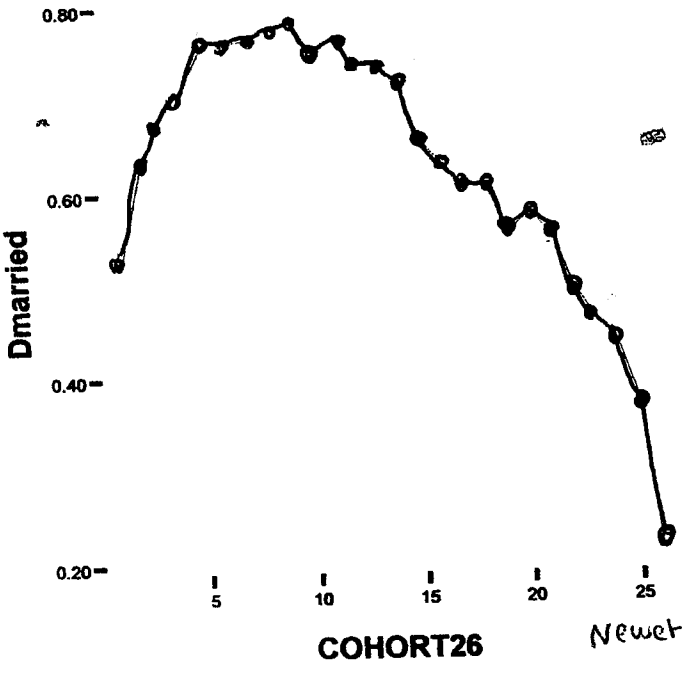
ABDEFECT

NO



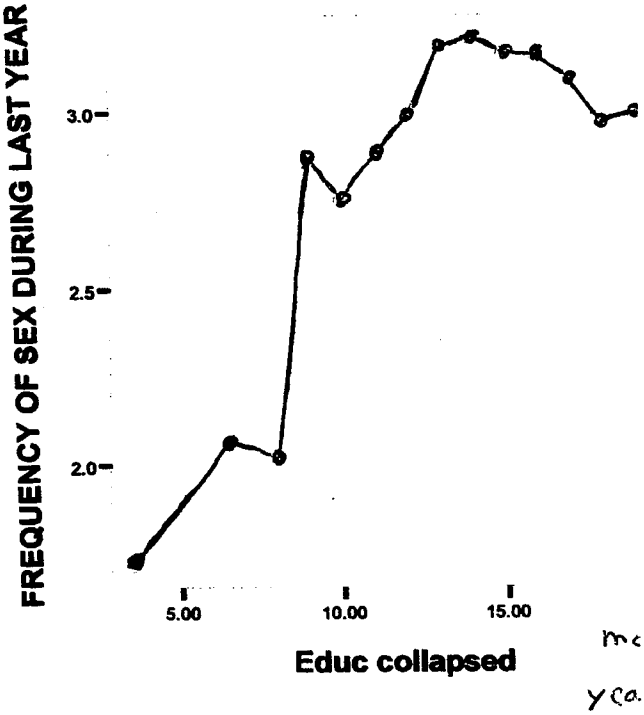
DMARRIED

NOW MARRIED



SEXFREQ

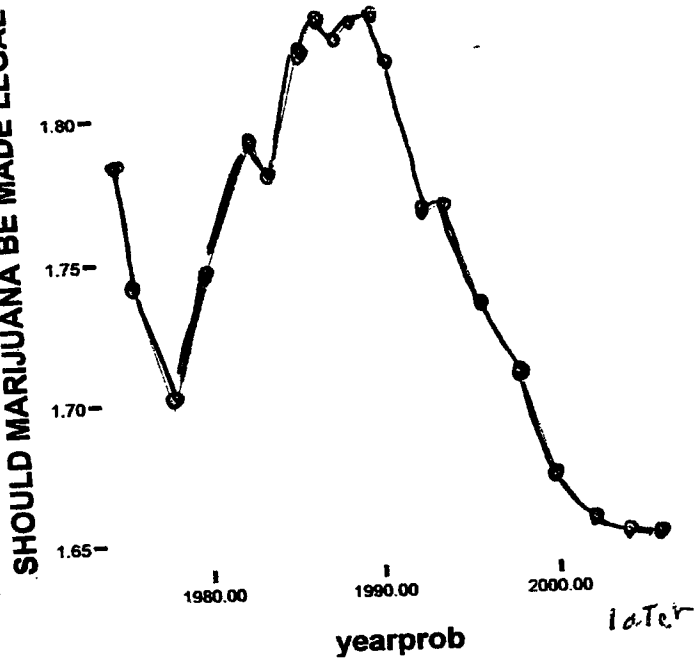
HIGH



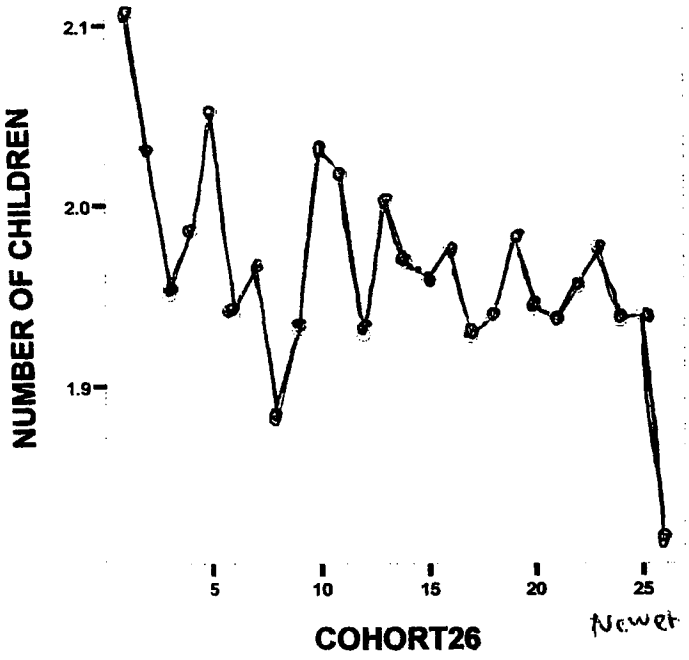
$r^2 = .30$ $r = .55$

HIGH SEXFREQ

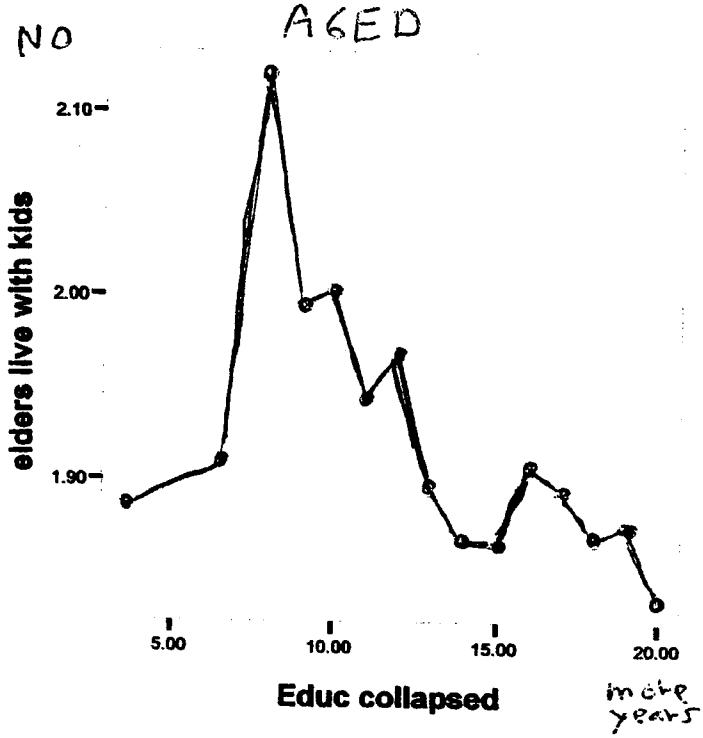
NO GRASS



MORE CHILDS

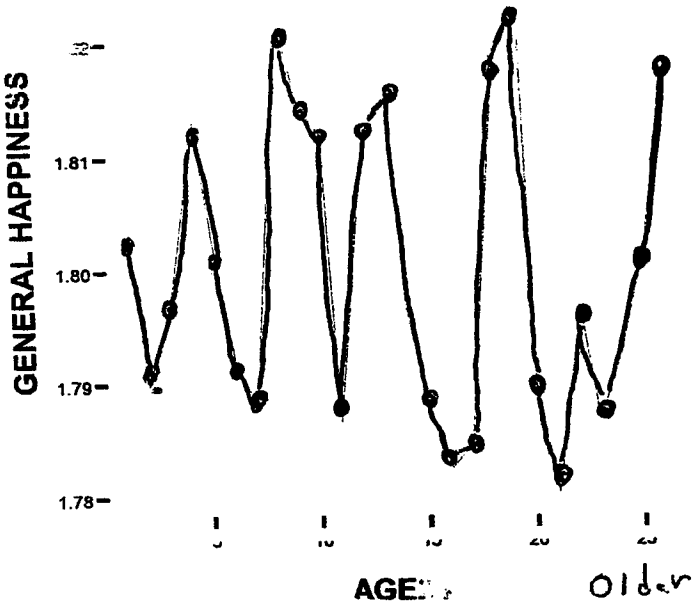


NO AGED



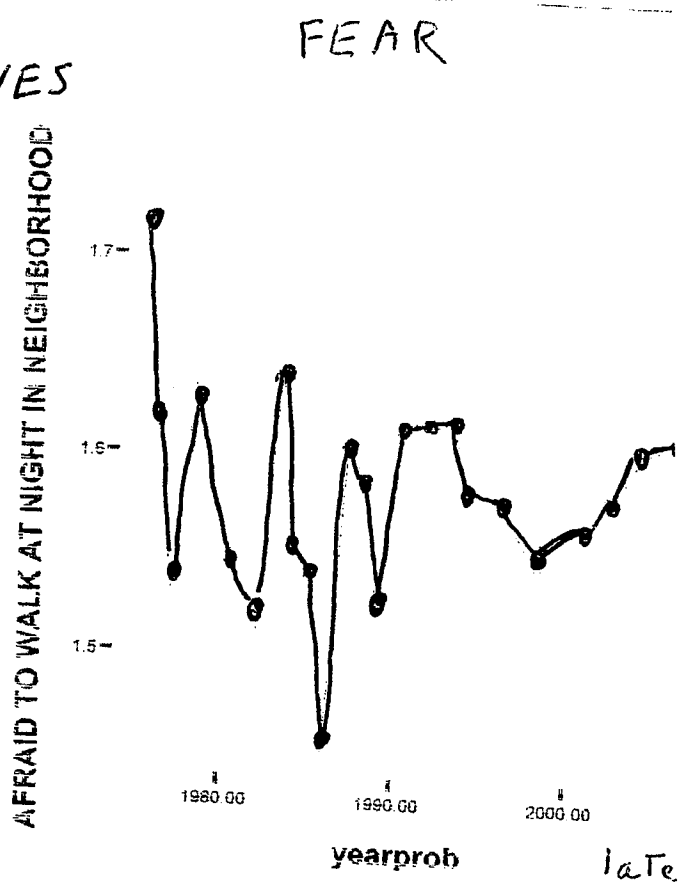
$r^2 = 0$ $r \leq .02$

NOT HAPPY

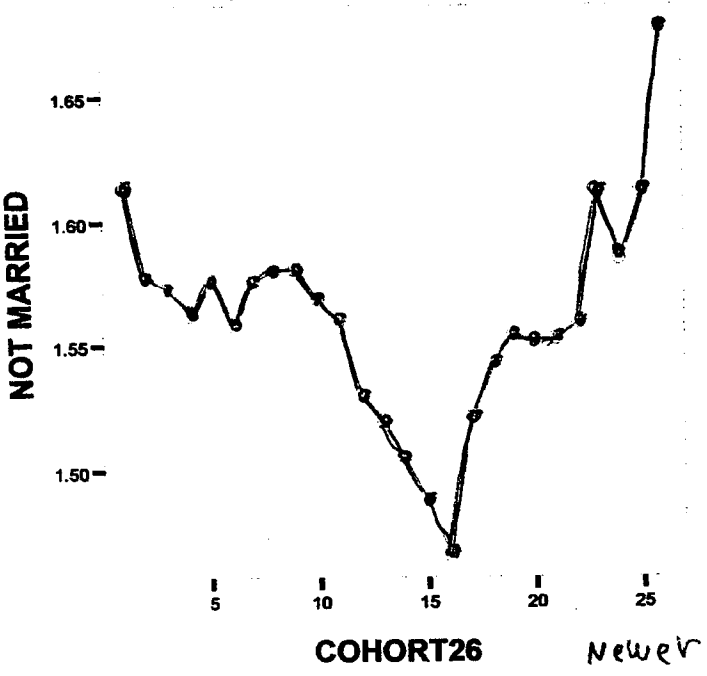


FEAR

YES



ABSINGLE



CAPPUN

OPPOSE

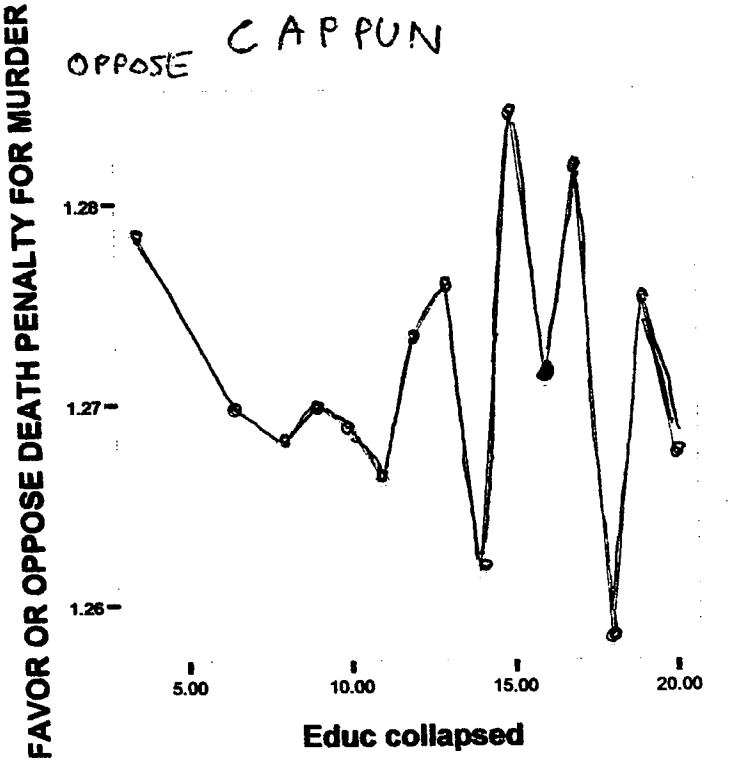
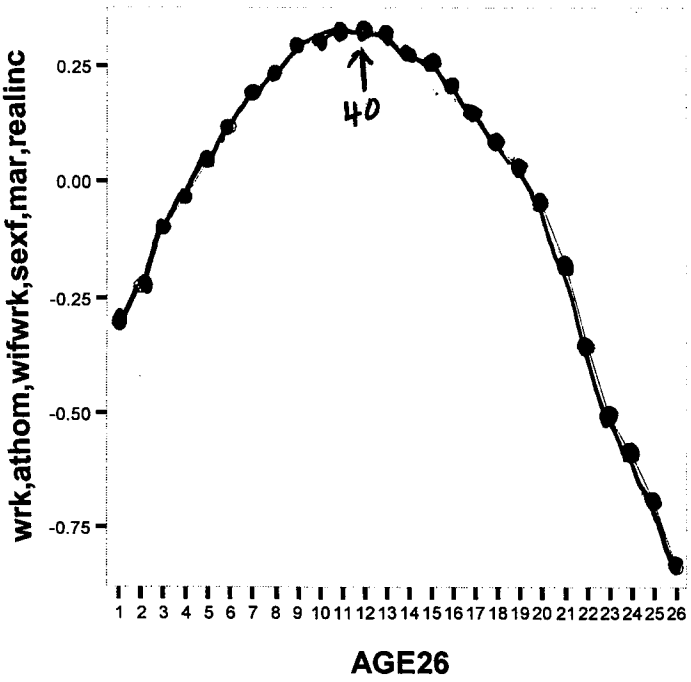


Figure 2

Life cycle Index by Age

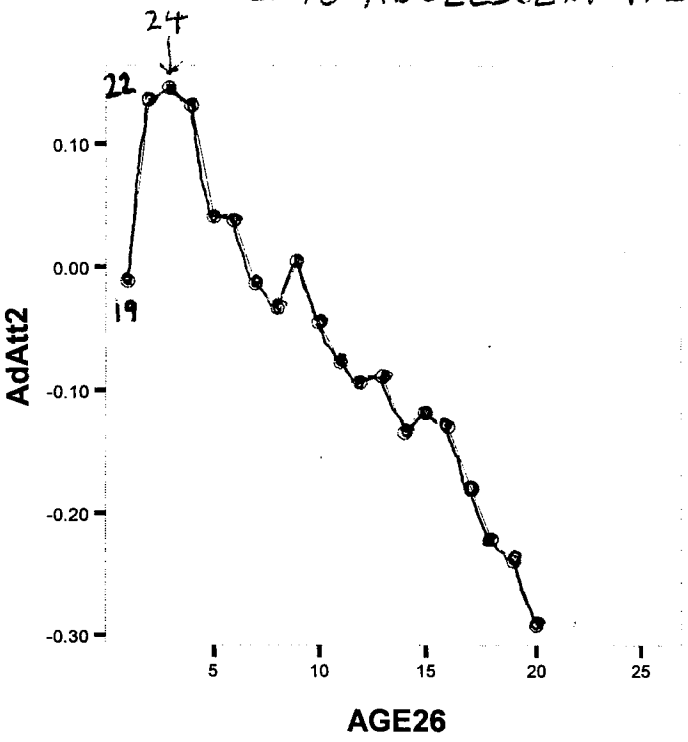


Dot/Lines show Means

N = 53,547

Figure 3

LATE ADOLESCENT ITEMS BY AGE



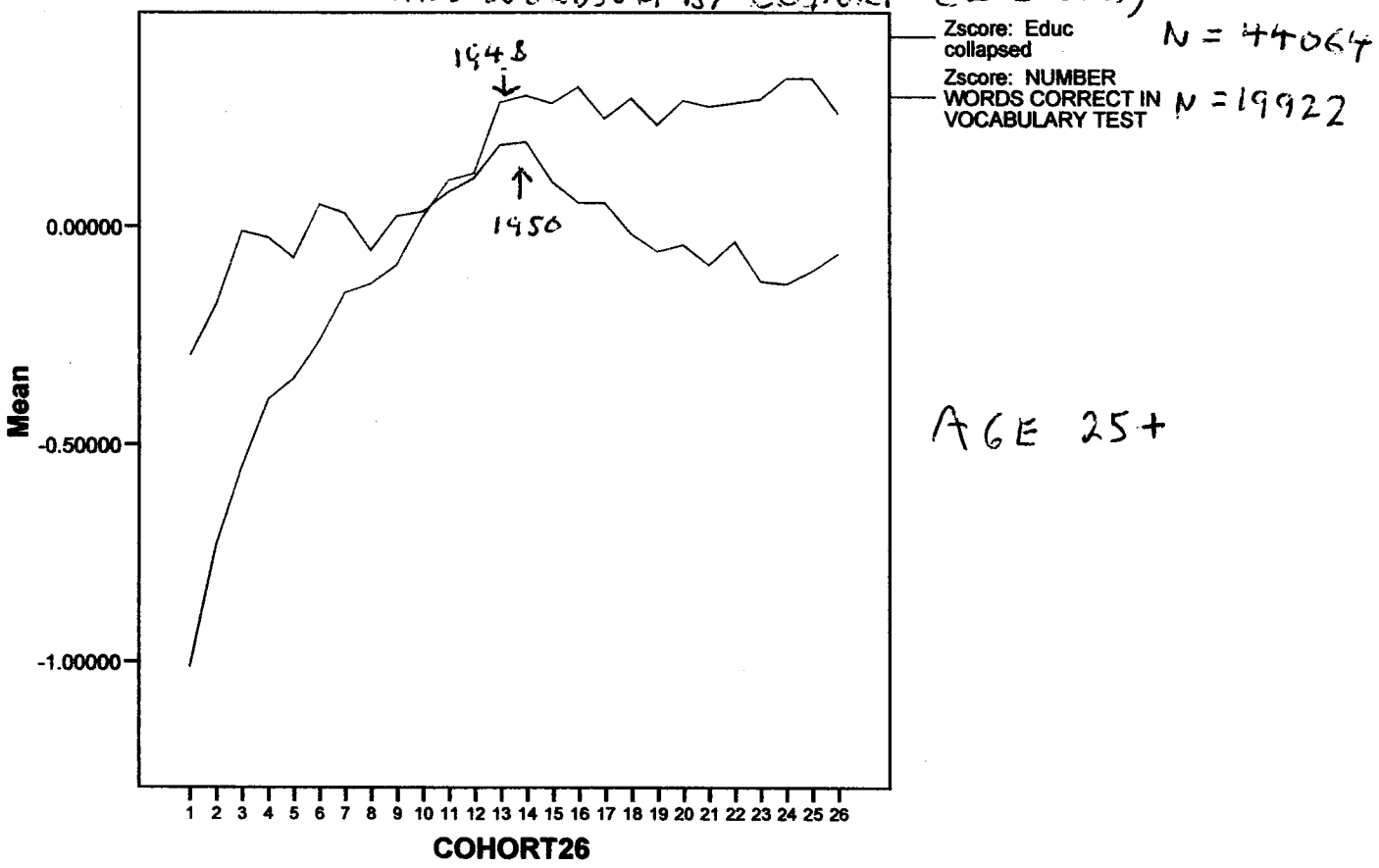
Dot/Lines show Means

N = 23,621

COHORT = 1947

Figur 4

EDUCATION AND WORDSUM BY COHORT (Z scores)



N = 44064

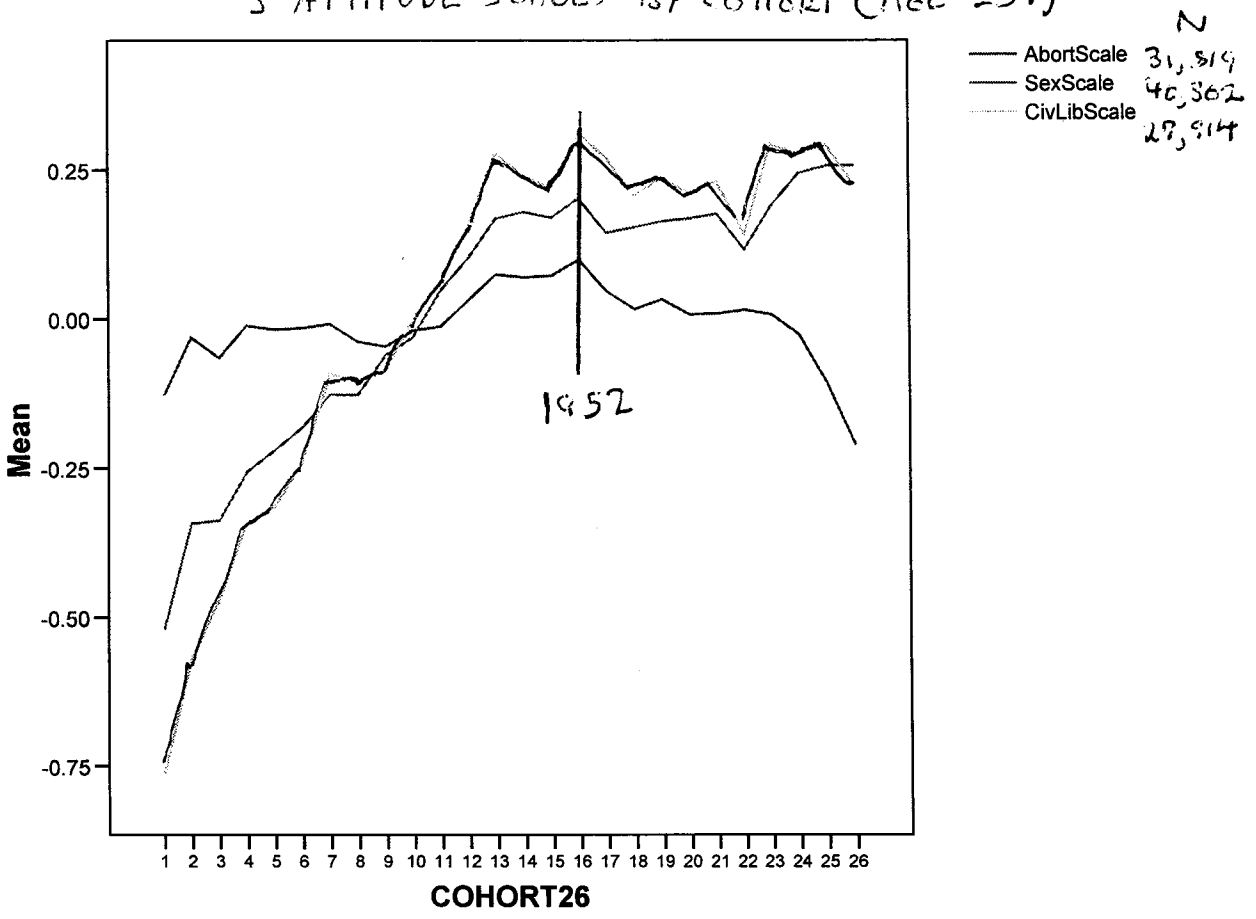
N = 19922

AGE 25+

Cases weighted by WAITER1

Figure 5

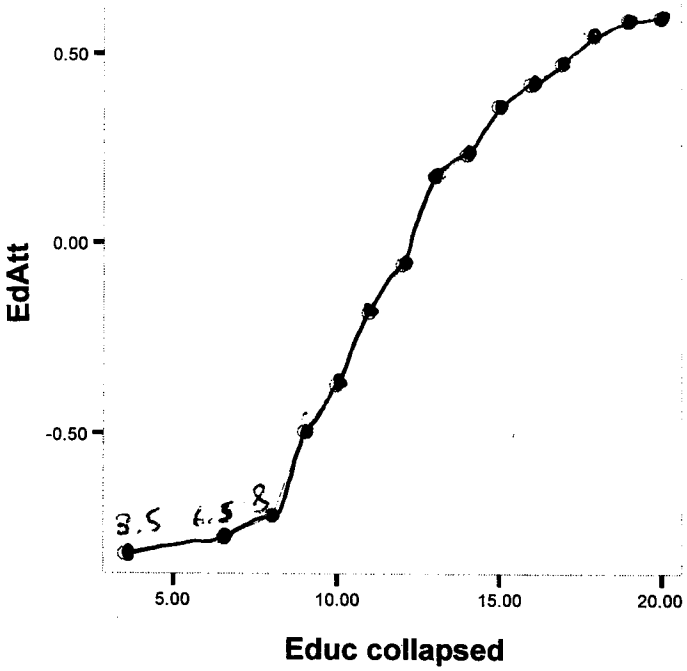
3 ATTITUDE SCALES BY COHORT (AGE 25+)



Cases weighted by WAITER1

Figure 6

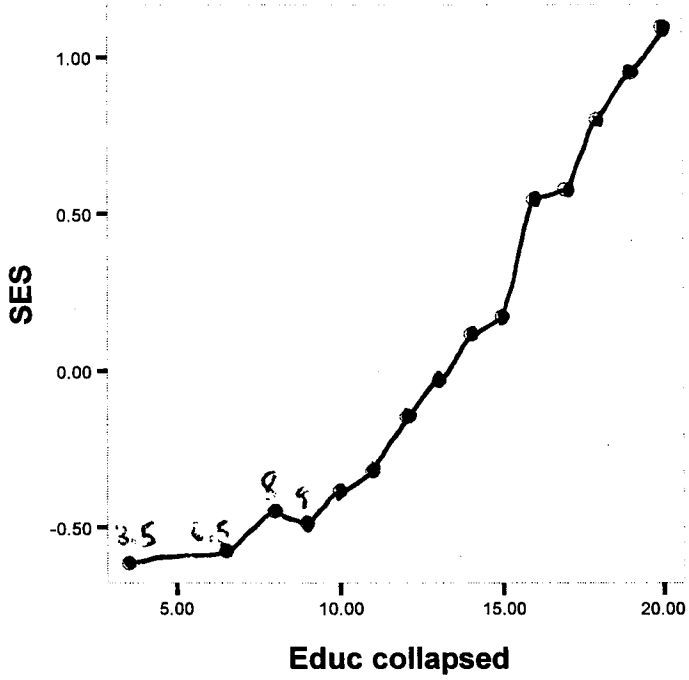
FIVE ATTITUDES BY EDUCATION



N = 37,660

Figure 7

Five SES items by education



N = 44,021