ON THE ORIGINS OF THE LATENT CURVE MODEL IN THE GROWTH CURVE AND FACTOR ANALYSIS TRADITIONS

Kenneth A. Bollen

University of North Carolina, Chapel Hill

Presented at the "Factor Analysis at 100: Historical Development and Future Directions" Conference. L.L. Thurstone Psychometric Laboratory, Department of Psychology, University of North Carolina, Chapel Hill, North Carolina, May 13-15, 2004.

OUTLINE OF PRESENTATION

I. INTRODUCTION

II. GENERIC MODEL

III. HISTORIC ROOTS OF LATENT CURVE MODELS (LCMs)

- A. Pre-1958: Individual Trajectories and Factor Analyses of Growth
- B. 1958: A Turning Point
- C. Issues of Case-by-Case & Factor Analysis Approaches
- D. Meredith and Tisak's (1984, 1990) Latent Curve Model

IV. LESSONS FOR CONTEMPORARY LCMs

V. CONCLUSIONS

I. INTRODUCTION

Growth in use & interest in LCMs

- availability of longitudinal data
- emphasis on individual differences
- accessibility of SEM software

Purposes:

1. Review & tribute to pioneers who laid foundation for LCMs
2. Draw lessons from historic work for contemporary LCMs

II. GENERIC MODEL

$$y_{it} = g_1(t)\eta_{i1} + g_2(t)\eta_{i2} + \cdots + g_K(t)\eta_{iK} + \varepsilon_{it}$$

y_{it} = repeated measure,

$g_k(t)$ = kth trajectory, function of t

η_{ik} = kth random coefficient or factor

$$E(\varepsilon_{it}) = 0 \qquad COV(\varepsilon_{it}, \eta_{ik}) = 0$$

$g_k(t)$ defines curves

η_{ik} defines the weight of case on curve

permits multiple curves underlying y_{it}

t time, age, trial, grade, etc.

Underidentified model

II. GENERIC MODEL

$$y_{it} = g_1(t)\eta_{i1} + g_2(t)\eta_{i2} + \cdots + g_K(t)\eta_{iK} + \varepsilon_{it}$$

Typical Latent Curve Model

$$y_{it} = \alpha_i + \lambda_t \beta_i + \varepsilon_{it}$$

Special case of Generic Model with:

$$K = 2$$

$$g_1(t) = 1 \qquad \eta_{i1} = \alpha_i$$

$$g_2(t) = \lambda_t = t - 1 \qquad \eta_{i2} = \beta_i$$

III. HISTORIC ROOTS OF LATENT CURVE MODELS (LCMs)

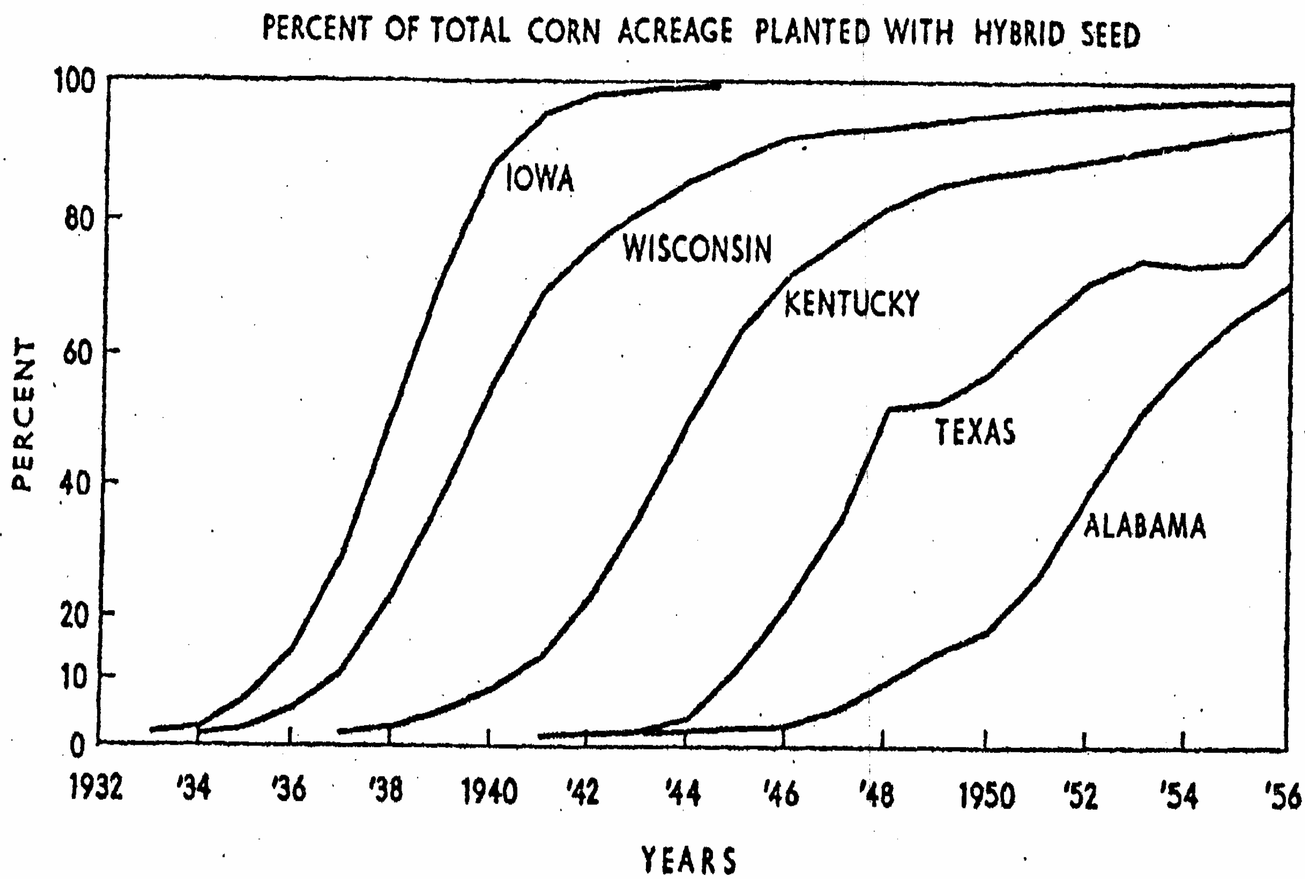
A. Pre-1958: Individual Trajectories and Factor Analyses of Growth

Wishart (1938)

- Weight of pigs over 17 times
- Quadratic function
- Individual rather than group
- Conditional growth model
- Case-by-case growth curves

Griliches (1957)

- Hybrid corn diffusion over states
- Logistic function
- Individual rather than group
- Conditional growth model
- Case-by-case growth curves



Source: Z. Griliches. (1957). "Hybrid corn: An Exploration in the Economics of Technological Change." Econometrica 25: 501-22.

III. HISTORIC ROOTS OF LATENT CURVE MODELS (LCMs)

A. Pre-1958: Individual Trajectories and Factor Analyses of Growth

Baker (1954)

Factor analysis of growth curves

Growth of peaches at 20 dates

3 Factor solution

no prespecified functional form

System estimator, not case-by-case

Underidentified model

Means & variances ignored

"The application of factor analysis to biological problems of growth seems to be novel so that one of the main purposes of this paper is to suggest such analyses" (1954, p.138).

In terms of earlier generic equation:

$$y_{it} = g_1(t)\eta_{i1} + g_2(t)\eta_{i2} + g_3(t)\eta_{i3} + \varepsilon_{it}$$

Rotated Factors			
F'_1	F'_2	F'_3	F'_4
.330	—.917	.127	.270
.511	—.893	—.140	—.293
.756	—.635	.177	—.112
.848	—.488	.193	.037
.936	—.253	.175	—.059
.960	—.056	.103	—.012
.957	—.041	.026	—.265
.982	.094	.065	—.072
.988	.084	.040	—.098
.986	.123	—.032	—.115
.990	—.002	—.046	—.139
.975	.051	.035	.133
.990	.010	.016	.032
.977	—.028	—.040	.054
.959	—.153	—.110	.321
.959	—.066	—.066	.132
.917	—.094	—.289	.005
.837	.048	—.501	.205
.718	.060	—.665	—.008
.681	.089	—.647	—.004

- "There is a central time of stability in the relative positions of the peaches extending from May 27, the date of pit hardening...until August 4. Sizeable loadings on this factor occur at all dates. The second factor is most clearly associated with early growth and since the loadings are negative, it is indicated that peaches relatively large at one early measurement tend to be relatively small at a near later measurement. ...Factor 2, F'_2 , complements factor 1 in the early stages of growth and fades away as factor 1 increases. Factor 3 complements factor 1 in the late stages of growth much the same as factor 1 does in the early stages. Factor 3 is again negative which probably indicates that the early ripening fruits complete growth earlier than some which ripen later which tends to reverse the relative positions of the fruits as they pass through time" Baker (1954, p. 141).

B. 1958: A TURNING POINT

Rao (1958)

Links growth curve models to factor analysis

$$y_{it}^{*} = g_1^{*}(t)\eta_{i1} + g_2^{*}(t)\eta_{i2} + \cdots + g_K^{*}(t)\eta_{iK} + \varepsilon_{it}^{*}$$

$$y_{it}^{*} = (y_{it} - y_{i,t-1})$$

$$g_k^{*}(t) = g_k(t) - g_k(t-1)$$

$$\varepsilon_{it}^{*} = (\varepsilon_{it} - \varepsilon_{i,t-1})$$

B. 1958: A TURNING POINT

Tucker (1958)

$$y_{it} = \phi(\theta_i, t)$$

approximate with

$$y_{it} = \sum_{k=1}^{K} f_k(t) h_k(\theta_i)$$

Define

$$g_k(t) \equiv f_k(t) \qquad \eta_{ik} \equiv h_k(\theta_i)$$

$$y_{it} = \sum_{k=1}^{K} g_k(t) \eta_{ik}$$

Matches generic equation except no error.

B. 1958: A TURNING POINT

Tucker's (1958) and Rao's (1958):

- Factor analysis as method to form flexible growth curves
- Factor loadings give shape of curve
- Factor values give weights for each case

BUT, underidentified model (rotation problem)

Tucker (1966)

powerful demonstration of approach

learning curves

- Subjects: enlisted army men
- Presentations of one of four letters
- Probabilities of letters: .7, .1, .1, .1
- 21 trials of 20 presentations
- # of times high prob. letter guessed
- 3 factor, orthogonal solution

TABLE 16-13

Scores on Rotated Factors
Group 70-10-10-10

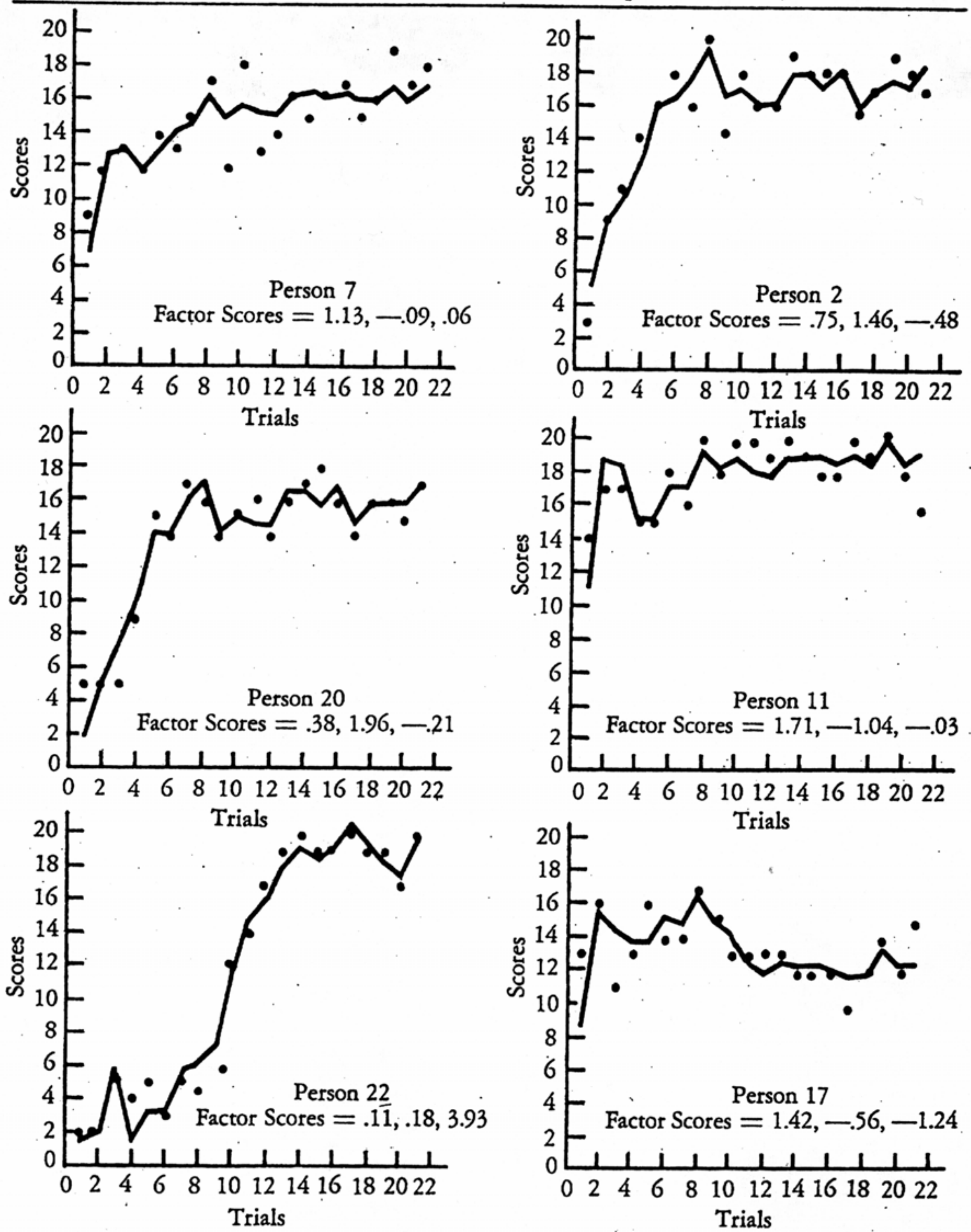
Individual	Rotated Factors *A*	*B*	*C*
1	1.16	−.70	.12
2	.75	1.46	−.48
3	.88	.04	.96
4	.52	1.48	.82
5	.84	−.04	.47
6	1.52	−.86	−.20
7	1.13	−.09	.06
8	1.25	−.86	.12
9	.61	.58	.51
10	1.40	−.57	−.30
11	1.71	−1.04	−.03
12	.81	.37	.75
13	.39	1.96	.32
14	.84	1.20	−.86
15	.56	.98	.92
16	.69	.47	.52
17	1.42	−.56	−1.24
18	.79	.78	−.59
19	.74	.99	−.91
20	.38	1.96	−.21
21	.58	1.72	−.76
22	.11	.18	3.93
23	1.34	−.63	.30
24	1.47	−.50	.36

Group 70-10-10-10

Factor *A*	Factor *B*	Factor *C*
6.17	−.23	.22
11.28	.51	.14
11.73	1.54	1.15
10.77	3.00	−.10
11.89	4.93	.21
12.86	4.72	.20
13.38	5.58	.83
14.87	5.94	.88
13.57	4.70	1.23
14.17	5.08	2.33
13.73	5.14	3.16
13.59	5.18	3.48
14.72	6.06	3.86
14.80	6.05	4.22
14.59	5.64	4.06
14.74	6.27	4.17
14.44	5.18	4.66
14.43	5.76	4.30
15.24	5.57	4.01
14.40	5.74	3.81
15.13	6.32	4.28

"These curves are generally near zero or positive and their slopes are either nearly zero or positive. ... Factors *B* and *C* start near zero while factor *A* is well above zero at all times. Factor *C* follows the base line for about six trials before starting to rise. All three curves seem to arrive at an asymptote by about trial 13 or 14. We might characterize *A* as early learning, *B* as middle learning, and *C* as late learning. The curve of factor *A* appears to have the shape of a negative exponential curve used extensively in learning theory while the curve of factor *C* appears to be more of the sigmoid shape which is also used in learning theory. Factor *B* lies somewhere between these two in shape" (Tucker, 1966, p. 495)

DIAGRAM 16-5. Selected Individual Learning Curves, Group 70-10-10-10



Source: L.R. Tucker. (1966) "Learning Theory and Multivariate Experiment: Illustration by Determination of Generalized Learning Curves." Pages 494-95.

C. ISSUES OF CASE-BY-CASE AND FACTOR ANALYSIS APPROACHES

1. Nature of disturbance/error

- no error (e.g., Tucker, 1958)

- ε_{it} as measurement error only (e.g., Rogosa, et al. 1982)

- most likely, error due to:
  - measurement error
  - approximation error
  - omitted variables
  - inherent stochastic component

C. ISSUES OF CASE-BY-CASE AND FACTOR ANALYSIS APPROACHES

2. Contrast of case-by-case and factor analysis approaches

case-by-case

- explicit functional form
- how to choose?
- fits individual cases
- collects parameter estimates

factor analysis

- estimates of functional forms through factor loadings
- system wide estimation
- factor score estimates as weights
- loadings & factors for indiv. predictions
- underidentified model

D. Meredith and Tisak's (1984, 1990) Latent Curve Model

Synthesis of :

1. Rao-Tucker on Exploratory Factor Analysis of growth curves and
2. Jöreskog's Confirmatory Factor Analysis methods

- Marks transition to contemporary factor analysis approach to growth curves.
- Identification conditions explored
  - Solves rotation problem
- Freed loading model
- Likelihood ratio tests
- Cohort sequential design
- Fit by conventional SEM software

IV. LESSONS FOR CONTEMPORARY LATENT CURVE MODELS

1. Emphasis on individual case rather than covariance matrix & means of LCMs
   – greater use of factor score predictions & factor loadings could correct this

2. Neglect of nonlinear trajectories in contemporary LCM practice
   – traditional analyses used nonlinear many LCM stick with linear

*corrections:

$g_k(t)$ specifications

estimated factor loadings

multiple factors

V. CONCLUSIONS

Basic ingredients for LCM

(1) approach that permits individual trajectories (e.g., Wishart, 1938; Griliches, 1957)

(2) ability to include prespecified linear or nonlinear trajectories (e.g., ditto)

(3) formulation as a factor analysis (e.g., Baker, 1954; Rao, 1958; Tucker, 1958)

(4) random trajectory coefficients as latent variables (same 3)

(5) ability to estimate shapes of curves or trajectories (same 3)

(6) identifying the model parameters (Meredith and Tisak, 1984; 1990)

(7) full incorporation of LCMs into SEMs (ditto)

V. CONCLUSIONS (cont)

LCMs development dependent on contributions from factor analysis and growth curve literature

Both traditional approaches still have lessons to teach to improve contemporary LCM practice