Xie, Yu. 1996 "A Demographic Approach to
Studying the Process of Becoming a
Scientist/Engineer" Pp.43-57 in *Careers in
Science and Technology: An International
Perspective*, edited by National Research
Council. Washington, D.C: National Academy
Press.

A Demographic Approach to Studying the Process of Becoming a Scientist/Engineer

Yu Xie

In this paper, I redefine career process as the collective experience of a birth cohort and propose a new demographic approach to studying the developmental process of becoming a scientist/engineer by following a synthetic birth cohort through its formative years of career development. The approach is dynamic rather than static, in the sense that it traces career changes over the life course of a cohort. At any given age, cohort members are identified as belonging to one of several states relevant to a scientific/engineering career. With data from longitudinal surveys, probabilities of cohort members' movements into and out of the different states are calculated as functions of age and population characteristics.

From these transitional probabilities, the process of becoming a scientist/engineer is modeled assuming a time-inhomogeneous Markov process commonly seen in standard multistate life tables. The time-inhomogeneity property of the Markov model makes an analysis adopting this approach non-parametric, descriptive, and capable of reproducing observed statistics with good data. Cross-sectional statistics on scientific careers can be generated from such a demographic model, for the size of any population is ultimately determined by inflow rates into and outflow rates out of the population.

The proposed approach has two important advantages over cross-sectional studies, which can only yield static snapshots of the population size with no information on the dynamic process of inflow and outflow. First, it makes it feasible to infer current and future descriptive statistics and other useful information on science/engineering (S/E) careers for any population or subpopulation with observed or hypothesized transition probabilities. Second, it allows the researcher to locate sources of attrition, especially for women and disadvantaged minorities, along the pipeline to becoming a scientist/engineer.

One major constraint for implementing the proposed approach in practical settings is the lack of longitudinal data, which are required for such dynamic analyses. When data from a true longitudinal design are unavailable, I propose that a synthetic cohort be constructed from different sources. For a demonstration of the new approach, I piece together data from two large data sets representing U.S. youth as they grow up between ages 13 and 32. The 1987-1991 Longitudinal Study of American Youth (LSAY) is used to obtain middle and high school students' (grades 7 through 12) interests in science education and changes of their interests over time. The 1972-1986 National Longitudinal Study of the Class of 1972 (NLS-72) is used to obtain information on years beyond high school, i.e., youth's probabilities of majoring in science, receiving science degrees at bachelor's and master's levels, and working in scientific occupations. Men and women are analyzed separately.

BASIC CONCEPTS AND METHODS

Cohort

In a classic article, Ryder (1965) defines a *cohort* as "the aggregate of individuals (within some population

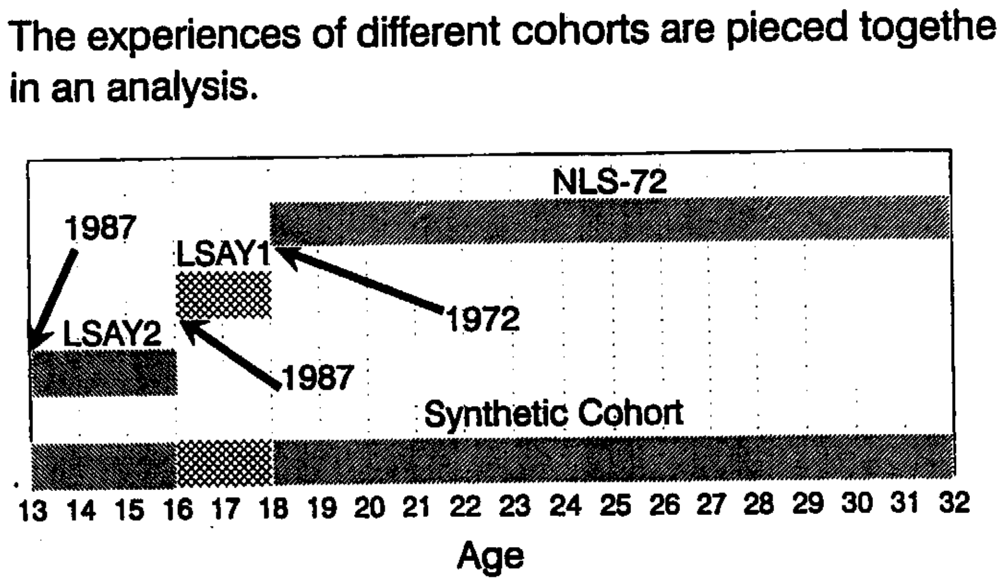

FIGURE 1 A synthetic cohort approach.

subject to scientific scrutiny.[2] For a given cohort, the career process coincides with maturation and aging, as in the case of an individual. Different from the individual-based definition, however, the cohort-based definition of career process allows the researcher to characterize the process using aggregate statistics with a certain degree of accuracy. Take the process of becoming a scientist/engineer as an example. Over time, some members of a birth cohort may stay in, move into, or move out of the S/E pool. Transition probabilities for these movements and non-movements are interesting characteristics of the cohort. In this paper, I explain how to study a cohort-based career process using these probabilities.

"Bathtub" Model of a Population

Any population can be described by the "bathtub" model: some people move in from outside while some people move out from inside. Simple as it appears, the "bathtub" model is actually a dynamic model. If the outflow rate exceeds the inflow rate for a sustained duration, the "bathtub" would eventually dry out; if the inflow rate exceeds the outflow rate for a sustained duration, the "bathtub" would eventually overflow.

Now consider the S/E pool as a population somehow-defined. For example, from early grades through high school, the population can be defined as students who are interested in science subjects and plan to attend college. In undergraduate and graduate years, the population can be defined as students majoring in science or obtaining science degrees. In the labor force, the population may be defined as workers in scientific occupations. As a birth cohort progresses through different stages, some cohort members move into the S/E pool while others move out of the pool. The task of a demographic analysis is to study the flows into and out of the pool and their implications for the process of becoming a scientist/engineer.

STATES AND TRANSITIONS

A *state* (denoted as s) is a distinct, well-defined, temporarily stable condition. A set of mutually exclusive and exhaustive states constitute a *state space* (denoted as S, $s \in S$, where $S = 1, \ldots M$). Let t denote time, which is assumed to be discrete in this paper (i.e., $t = 1, \ldots T$).

Transition probability, $p_{ij}(v, w)$, is the probability that an individual will be in state j at $t=w$ given he or she is in state i at $t=v$ ($v < w$):

$$p_{ij}(v,w) = Prob(s=j, t=w \mid s=i, t=v).$$

Let $P(v, w)$ be a square matrix that contains all elements of $p_{ij}(v, w)$:

second is the 1972-1986 NLS-72. In LSAY two high school cohorts were followed up every semester, one from grade 7 in 1987 to grade 10 in 1990 and the other from grade 10 in 1987 to grade 12 in 1989. The NLS-72 cohort was grade 12 in 1972 and was followed up in 1973, 1974, 1976, 1979, and 1986.[4] I treat the three cohorts as if they were part of a single cohort. The surveys provide enough information to cover the hypothetical cohort continuously from ages 13 to 32.[5]

The pitfall of this research strategy is, of course, that the experiences of the synthetic cohort do not represent those of any real cohort. In this paper, the data sets used might be problematic, as the earlier years of the synthetic cohort were observed much later (starting in 1987) than the later years of the synthetic cohort (starting in 1972). Without better data, my analysis of the life course process of becoming a scientist/engineer assumes that career process is relatively stable, i.e., age-dependent rather than cohort- or period-dependent. This assumption is consistent with an observation made more than 40 years ago by Ginzberg and his associates (Ginzberg et al., 1951) that the career process is a developmental process, thus, an age-dependent process.

Each fall, the LSAY survey asked high school students whether they would "enroll in a four-year college or university" upon graduation. I consider those students who answered "yes" as intending to obtain postsecondary education. The same questionnaire also asked whether students agree or disagree with the statement "I enjoy science" as follows: "strongly agree," "agree," "not sure," "disagree," and "strongly disagree." I classify students who responded "strongly agree" *and* planned to enroll in a four-year college or university as belonging to the state of "intended science/engineering postsecondary education." This measure is constructed for all high school years of both cohorts in LSAY.

In each follow-up, NLS-72 respondents were asked to report their actual or intended fields of study if they were attending postsecondary school. The answers were coded into six-digit numerical codes. From the 1972 base year to the 1979 fourth follow-up, NLS-72 used a coding system described by *A Taxonomy of Instructional Programs in Higher Education*, commonly referred to as HEGIS (Huff and Chandler, 1970). For the fifth follow-up, NLS-72 changed to a new system described by *A Classification of Instructional Programs* (Malitz, 1981). From the two systems of instructional programs, I extracted a number of detailed codes as S/E fields of study.[6] In 1976 and 1979, NLS-72 ascertained whether respondents had obtained bachelor's and master's degrees. For those with positive responses, NLS-72 ascertained detailed fields of their degrees. In the 1986 fifth follow-up, NLS-72 changed the question and collected information regarding respondents' highest degrees (from baccalaureates to doctorates); data on fields of degrees, however, were not collected. To solve this problem, auxiliary information is used with additional assumptions. Respondents who undertook postsecondary education between 1979 and 1986 were asked to report their latest fields of study. Thus, for a respondent who obtained additional postsecondary education and whose highest degree in 1986 was more advanced than or equal to his or her highest degree in 1979, his or her latest field of study is imputed as the field of his or her highest degree.

I hereby create the following seven educational states:

(0) secondary education or college dropout
(1) intended or actual non-S/E postsecondary education
(2) intended or actual S/E postsecondary education
(3) non-S/E bachelor's degree
(4) S/E bachelor's degree
(5) non-S/E master's degree
(6) S/E master's degree

Figure 2 presents schematic flows among the seven educational states; the solid lines represent typical flows and the dotted lines represent untypical flows.[7]

From the above operational definitions of the states, I calculate, separately for males and females, 13 transition probability matrices for the hypothetical cohort, given in Tables A1 and A2. The reported rates and counts were appropriately weighted to reflect the sampling designs and non-responses so that resulting transition rates are the best estimates of their corresponding conditional probabilities in the population.[8] These transition rates are mainly used in analysis to be reported later. It appears that the use of different measures and different data sets may bias the empirical results. One way to check for potential biases is to compare marginal distributions at ages 16 and 18 that connect two different cohorts (see Figure 1) since these marginal distributions are observed twice. The comparison gives acceptable results (could be derived from Tables A1 and A2). It should also be noted that in the empirical analysis to be reported later, for all ages except the initial condition (age 13), only *transi-*

TABLE 1 Distribution of Educational States by Age and Sex for a Synthetic Cohort

		Educational State							
Age	School Age	0	1	2	3	4	5	6	Total
Panel A: Females									
13	Grade 7	320.0	514.0	166.0	0.0	0.0	0.0	0.0	1000
14	Grade 8	363.3	498.6	138.2	0.0	0.0	0.0	0.0	1000
15	Grade 9	311.8	544.6	143.7	0.0	0.0	0.0	0.0	1000
16	Grade 10	321.2	545.6	133.2	0.0	0.0	0.0	0.0	1000
17	Grade 11	406.0	476.5	117.5	0.0	0.0	0.0	0.0	1000
18	Grade 12	478.0	469.3	52.7	0.0	0.0	0.0	0.0	1000
19	College 1	445.0	518.7	36.3	0.0	0.0	0.0	0.0	1000
20	College 2	541.0	429.1	29.9	0.0	0.0	0.0	0.0	1000
21	College 3	617.6	352.3	30.2	0.0	0.0	0.0	0.0	1000
22	College 4	705.3	268.8	26.0	0.0	0.0	0.0	0.0	1000
23	Postgraduate 1	740.2	83.7	7.0	152.6	16.5	0.0	0.0	1000
26	Postgraduate 4	706.4	69.5	5.6	172.4	22.4	21.6	2.1	1000
32	Postgraduate 11	564.8	174.0	10.3	169.4	18.3	57.7	5.4	1000
Panel B: Males									
13	Grade 7	380.0	425.0	195.0	0.0	0.0	0.0	0.0	1000
14	Grade 8	422.0	382.0	196.0	0.0	0.0	0.0	0.0	1000
15	Grade 9	365.8	478.1	156.1	0.0	0.0	0.0	0.0	1000
16	Grade 10	420.2	454.5	125.3	0.0	0.0	0.0	0.0	1000
17	Grade 11	480.5	394.5	125.0	0.0	0.0	0.0	0.0	1000
18	Grade 12	520.4	380.9	98.7	0.0	0.0	0.0	0.0	1000
19	College 1	479.2	418.2	102.6	0.0	0.0	0.0	0.0	1000
20	College 2	532.5	375.4	92.1	0.0	0.0	0.0	0.0	1000
21	College 3	582.9	320.4	96.7	0.0	0.0	0.0	0.0	1000
22	College 4	681.9	248.4	69.7	0.0	0.0	0.0	0.0	1000
23	Postgraduate 1	739.3	88.3	19.5	111.5	41.3	0.0	0.0	1000
26	Postgraduate 4	692.5	68.1	20.8	140.4	53.5	17.9	6.9	1000
32	Postgraduate 11	569.3	136.0	35.6	141.1	40.5	65.5	12.0	1000

NOTE: Educational states are defined as: (0) secondary education or college dropout, (1) intended or actual non-S/E postsecondary education, (2) intended or actual S/E postsecondary education, (3) non-S/E bachelor's degree, (4) S/E bachelor's degree, (5) non-S/E master's degree, and (6) S/E master's degree.

TABLE 2 Number of Females in Science/Engineering Educational States at Age 32 if Male's Transition Rates Were True for Females at Different Ages

Males' Rates were Used at Age	School Age	Educational State 2	% Explained Gender Gap	4	% Explained Gender Gap	6	% Explained Gender Gap
No Substitution (Observed)		10.3		18.3		5.4	
13	Grade 7	10.3	0.0%	18.3	0.0%	5.4	0.0%
14	Grade 8	10.3	0.0	18.3	-0.2	5.4	-0.2
15	Grade 9	10.3	0.1	18.2	-0.8	5.4	-0.9
16	Grade 10	10.3	0.0	18.4	0.1	5.4	-0.1
17	Grade 11	10.3	0.0	19.0	2.9	5.6	1.7
18	Grade 12	10.3	-0.1	20.2	8.4	5.8	4.9
19	College 1	10.2	-0.3	19.9	6.9	5.8	5.1
20	College 2	10.2	-0.5	22.9	20.7	6.3	13.4
21	College 3	10.3	-0.1	20.9	11.8	5.9	7.0
22	College 4	10.7	1.7	21.1	12.6	5.3	-1.7
23	Postgraduate 1	11.8	5.8	24.1	25.8	6.8	21.5
26	Postgraduate 4	33.4	91.0	20.6	10.4	6.8	21.0

NOTE: Educational states are defined as: (0) secondary education or college dropout, (1) intended or actual non-S/E postsecondary education, (2) intended or actual S/E postsecondary education, (3) non-S/E bachelor's degree, (4) S/E bachelor's degree, (5) non-S/E master's degree, and (6) S/E master's degree. The exercise alternately substitutes one set of males' transition rates at a given age while keeping everything else intact.

Exit and Entry Rates

Given our simple "bathtub" model, women's underrepresentation in S/E educational states could result from two sources: women's exit rate from S/E states is higher than men's or women's entry rate into S/E states is lower than men's. Two series of exit and entry rates are displayed in Figures 5 and 6.[11] In calculating the rates, I combined the three S/E states into a single state and lumped the four non-S/E states into another state. An interesting result is that the exit rate for females trails that for males closely except at age 17. A larger and more consistent gender gap, however, is observed for the entry rate after age 17. From these two figures, I infer that a large portion of the gender gap in attaining S/E education is not merely due to women's higher likelihood to exit the S/E pool. Men are just as likely as women to drop out of the S/E pool, but their likelihood to enter or re-enter the pool once out of it is significantly higher, particularly in later years. Unfortunately, past research has not paid attention to this problem. For example, my results contradict Berryman's (1983, p.7) assertion that "*after high school, migration is almost entirely out of, not into, the pool*" (emphasis original). For males, I have found that the rate of migration into the pool is around 4 percent, compared to less than 2 percent for women.

As small as they may seem, cumulatively these figures are very significant given that only 2 percent of females and 5 percent of males obtain bachelor's and master's degrees in science by age 32 (see Table 1).

S/E Occupations

Education affects occupation, but only in non-deterministic ways. Obtaining S/E education means that one's likelihood of working in a S/E occupation significantly increases, but it cannot be equated with S/E occupation. In fact, Table 4 shows that only 9 percent of females and 35 percent of males with S/E bachelor's degrees have S/E occupations at age 32. At the master's level, the percentages are better, 27 percent for females and 54 percent for males.

In Table 5, I present simulated occupational distributions at age 32 under four different conditions.[12]

In Panel A, females' educational distribution at age 32 (last line, Panel A of Table 4) is used. In Panel B, males' educational distribution at age 32 (last line, Panel B of Table 4) is used. Within each panel of Table 5, two lines represent two sets of transition rates from educational states to occupational states, one for females and one for males. Thus, the first line of Panel A and the second line of Panel B are simulated distributions respectively for females and males using gender-specific information. That is, for a 1,000-member female cohort following the age-specific transition rates observed for our data, only 14.5 of them work as scientists/engineers at age 32. The comparable number is 59.5 for a 1,000-member male cohort. Most of the gender gap is due to women's lower likelihood to work as scientists/engineers given the same educational background. If females had the same distribution of educational states as males, female scientists/engineers would increase to 17.5 per 1,000 (first line of Panel B). However, if females had the same Table 4 transition rates from educational states to occupational states at rates of transition from education to occupation as males, female scientists/engineers would increase to 49.8 per 1,000. Therefore, women's lower achievement in attaining S/E education can only explain a small fraction (about 10 percent) of the gender gap in attaining S/E occupations.

CONCLUSION

In brief, I offer the following conclusions:

1. This paper proposes a new demographic approach to studying the process of becoming a scientist/engineer.
2. The proposed approach consists of constructing a synthetic cohort from different longitudinal surveys and modeling the career process as a Markov process as the cohort ages (or matures).
3. An age pyramid in the S/E educational pool is found to exist. Generally speaking, for any given cohort, the proportion of people remaining in the pool decreases with age.
4. Women's representation in the S/E educational pool drops suddenly near high school graduation (between ages 17 and 18).
5. College and graduate years account for most of the sex differences in the proportion attaining science degrees by early adulthood.
6. Men and women differ more in the entry or re-entry rate into the S/E educational pool than the exit rate out of the S/E educational pool.
7. Women's underrepresentation in S/E occupations is mainly due to women's lower likelihood of being employed in S/E occupations net of differential access to S/E education rather than women's lower likelihood of having a S/E education.

TABLE 4 Transition Rates From Educational States to Occupational States at Age 32 by Sex

	Occupational States			
Educational State	Not Working	Non S/E	S/E	(n)
Panel A: Females				
Secondary education only	34.41%	65.09	0.50	(3583)
Non-S/E postsecondary education	25.32	72.61	2.07	(1110)
S/E postsecondary education	16.67	81.67	1.67	(60)
Non-S/E bachelor's degree	25.92	71.84	2.24	(1115)
S/E bachelor's degree	35.34	55.64	9.02	(133)
Non-S/E master's degree	16.16	82.17	1.67	(359)
S/E master's degree	36.36	36.36	27.27	(33)
Panel B: Males				
Secondary education only	12.64%	84.97	2.39	(3354)
Non-S/E postsecondary education	15.14	79.63	5.22	(766)
S/E postsecondary education	15.67	77.42	6.91	(217)
Non-S/E bachelor's degree	4.57	87.45	7.98	(940)
S/E bachelor's degree	6.38	57.38	36.24	(298)
Non-S/E master's degree	7.13	86.94	5.94	(421)
S/E master's degree	5.06	40.51	54.43	(79)

NOTE: Main entries are row percentages, i.e., estimated probability of occupational states conditional on educational states.

TABLE A1 Matrices of Transition Among Educational States, Females

	Origin				Destination State				
School Ages	State	0	1	2	3	4	5	6	(n)
Grade 7 to Grade 8	0	0.6219	0.3060	0.0721	--	--	--	--	(402)
	1	0.2407	0.6728	0.0864	--	--	--	--	(648)
	2	0.2440	0.3301	0.4258	--	--	--	--	(209)
Grade 8 to Grade 9	0	0.5718	0.3536	0.0746	--	--	--	--	(362)
	1	0.1676	0.7156	0.1168	--	--	--	--	(531)
	2	0.1481	0.4296	0.4222	--	--	--	--	(135)
Grade 9 to Grade 10	0	0.6532	0.2626	0.0842	--	--	--	--	(297)
	1	0.1598	0.7648	0.0754	--	--	--	--	(557)
	2	0.2123	0.3288	0.4589	--	--	--	--	(146)
Grade 10 to Grade 11	0	0.8220	0.1602	0.0178	--	--	--	--	(337)
	1	0.2313	0.6886	0.0801	--	--	--	--	(562)
	2	0.1185	0.3704	0.5111	--	--	--	--	(135)
Grade 11 to Grade 12	0	0.8410	0.1503	0.0087	--	--	--	--	(346)
	1	0.2450	0.7275	0.0275	--	--	--	--	(400)
	2	0.1683	0.5248	0.3069	--	--	--	--	(101)
Grade 12 to College 1	0	0.7767	0.2140	0.0093	--	--	--	--	(3883)
	1	0.1455	0.8332	0.0213	--	--	--	--	(3196)
	2	0.1034	0.4803	0.4163	--	--	--	--	(406)
College 1 to College 2	0	0.9000	0.0963	0.0037	--	--	--	--	(5069)
	1	0.2576	0.7300	0.0124	--	--	--	--	(5012)
	2	0.1910	0.2095	0.5995	--	--	--	--	(377)
College 2 to College 3	0	0.8966	0.0976	0.0058	--	--	--	--	(5715)
	1	0.2973	0.6790	0.0237	--	--	--	--	(4016)
	2	0.1644	0.2705	0.5651	--	--	--	--	(292)
College 3 to College 4	0	0.9285	0.0666	0.0048	--	--	--	--	(5988)
	1	0.3506	0.6291	0.0203	--	--	--	--	(2313)
	2	0.2757	0.2000	0.5243	--	--	--	--	(185)
College 4 to Postgraduate 1	0	0.9036	0.0757	0.0079	0.0119	0.0009	--	--	(6565)
	1	0.3517	0.1052	0.0036	0.5322	0.0072	--	--	(1939)
	2	0.3239	0.0795	0.0170	0.0455	0.5341	--	--	(176)
Postgraduate 1 to Postgraduate 4	0	0.8992	0.0766	0.0057	0.0161	0.0015	0.0008	--	(6093)
	1	0.4523	0.1454	0.0088	0.3604	0.0183	0.0142	0.0007	(1479)
	2	0.4179	0.0821	0.0896	0.0597	0.3358	0.0149	0.0000	(134)
	3	--	--	--	0.8387	0.0335	0.1254	0.0024	(1252)
	4	--	--	--	0.1159	0.7464	0.0362	0.1014	(138)
Postgraduate 4 to Postgraduate 11	0	0.7712	0.1992	0.0097	0.0169	0.0013	0.0018	0.0000	(3916)
	1	0.2689	0.4499	0.0391	0.2029	0.0122	0.0220	0.0049	(409)
	2	0.2308	0.3590	0.1282	0.1795	0.0513	0.0256	0.0256	(39)
	3	--	--	--	0.7990	0.0174	0.1732	0.0104	(1149)
	4	--	--	--	0.2098	0.5944	0.1608	0.0350	(143)
	5	--	--	--	--	--	0.9426	0.0574	(122)
	6	--	--	--	--	--	0.4545	0.5455	(11)

NOTE: Main entries are row proportions. Educational states are defined as: (0) secondary education only, (1) intended or actual non-S/E postsecondary education, (2) intended or actual S/E postsecondary education, (3) non-S/E bachelor's degree, (4) S/E bachelor's degree, (5) non-S/E master's degree, and (6) S/E master's degree. Cells omitted and marked with "--" are structural zeros.

REFERENCES

Bartholomew, D. J. 1973. Stochastic Models for Social Processes, 2nd edition. London: Wiley.

Berryman, S. E. 1983. Who Will Do Science? A special Report, the Rockefeller Foundation.

Brown, D., L. Brooks and Associates (eds). 1990. Career Choice and Development, 2nd edition. San Francisco: Jossey-Bass Publishers.

Elder, G. H., Jr. 1985. Perspectives on the Life Course. Pages 23-49 in Life Course Dynamics: Trajectories and Transitions, 1968-1980, edited by Glen H. Elder, Jr. Ithaca, NY: Cornell University Press.

Freeman, R. B. 1971. The Market for College-Trained Manpower: A Study in the Economics of Career Choice. Cambridge, MA: Harvard University Press.

Ginzberg, E., S. W. Ginsburg, S. Axelrad, and J.L. Herma. 1951. Occupational Choice: An Approach to a General Theory. New York: Columbia University Press.

Huff, R. A. and M. O. Chandler. 1970. A Taxonomy of Instructional Programs in Higher Education. Washington, D.C.: U.S. Printing Office.

Hoem, J. M. 1977. A Markov Chain Model of Working Life Tables. Scandinavian Actuarial Journal, 1-20.

Land, K. C. and G. C. Hough, Jr. 1989. New Methods for Tables of School Life, With Applications to U.S. Data from Recent School Years. Journal of the American Statistical Association, 84:63-75.

Malitz, G. S. 1981. A Classification of Instructional Programs. Washington, D.C.: U.S. Printing Office.

Namboodiri, K. and C. M. Suchindran. 1987. Life Table Techniques and their Applications. Orlando, FL: Academic Press.

National Science Board. 1986. Undergraduate Science, Mathematics and Engineering Education. Report of the NSB Task Committee on Undergraduate Science and Engineering Education.

Quetelet, L. A. J. 1842. A Treatise on Man and the Development of his Faculties, a facsimile reproduction of the English translation of 1942. Gainesville, FL: Scholar's Facsimiles and Reprints, 1969.

Ryder, N. B. 1965. The Cohort as a Concept in the Study of Social Change. American Sociological Review, 30:843-861.

Sewell, W. H., A. O. Haller, and A. Portes. 1969. The Educational and Early Occupational Attainment Process. American Sociological Review, 34:82-92.

Tuma, N. B. and M. T. Hannan. 1984. Social Dynamics: Models and Methods. San Francisco: Academic Press.

Xie, Y. 1989. The Process of Becoming a Scientist. Unpublished Dissertation, University of Wisconsin.