CODEBOOK SUPPLEMENT

APPENDIX #9

The material which follows originated in Determinants of Early Labor Market Success Among Young Men: Race, Ability, Quantity and Quality of Schooling, Andrew I. Kohen, Ph. D. dissertation, January 1973. The discussion and results reported here, though based on the National Longitudinal Surveys of Young Men 14-24, apply equally well to the Surveys of Young Women 14-24.

APPENDIX A

ON THE POOLING OF MENTAL ABILITY MEASURES FROM DIFFERENT TESTS: A PRAGMATIC APPROACH

by

Robert E. Herriott and Andrew I. Kohen*

For some time there has existed within the psychometric literature a general skepticism regarding the usefulness of pooling measures of mental ability obtained from different tests. In particular, it is urged (if not insisted) that investigators seeking to measure mental ability administer the same test to all subjects under the same highly standardized conditions. Yet, there are many instances in large scale social research on geographically dispersed samples where such uniformity in data collection procedures is not possible. Under such circumstances should the investigator abandon his theoretical interest in mental ability, or should he proceed in a more pragmatic fashion?

Recently a unique opportunity arose for examining empirically the consequences of pooling data from a large number of different tests of mental ability. As one part of a National Longitudinal Survey of Young Men[1] the U.S. Bureau of the Census sent inquiries to 2,042 secondary

---

*This appendix is a revised version of an earlier, unpublished paper by the authors. Professor Herriott is Director of the Center for the Study of Education, Institute for Social Research, The Florida State University. The authors wish to acknowledge the assistance of John Grasso and Martin Mehall in the computer processing of the data.

[1]This group constitutes one of the four population samples comprising the National Longitudinal Studies (LGS) being carried out by The Ohio

schools[2] to obtain the most recent data regarding the mental ability of males between the ages of 14 and 24 who either were currently attending that school or who had most recently attended it. Through extensive follow-ups involving both remailings and long-distance telephone calls, scores obtained from over 30 different tests of mental ability were received for 3,375 of the 4,007 males for whom scores were sought.[3] Presented below is a review of some of the psychometric issues underlying the equating of scores from different tests, as well as a description of procedures used to transform the available scores, and to assess their comparability. In addition some suggestions for improving the quality of this type of data are offered.

Psychometric Issues

In the psychometric literature an important distinction is made between tests of the same "function" and tests of different functions. Tests of the same function are said to be "parallel" and those of different functions "non-parallel." Although the definition of function is not always clear-cut, it is generally assumed that alternate forms of the

---

State University Center for Human Resource Research under a contract with the Manpower Administration of the U.S. Department of Labor.

[2]Actually, the survey of secondary schools contained the 3,030 institutions attended by members of two samples of 14 to 24 year olds, i.e., males and females. However, members of the male sample attended only 2,042 of the schools; the remaining 988 schools were represented in the samples only by females. Many of the 2,042 schools had pupils in both sex cohorts. The school survey instrument appears in Appendix G.

[3]The LGS sample of male youth initially consisted of 5,225 respondents, but scores were sought only for the 4,007 young men who (1) had completed the ninth grade by the time of the survey and (2) had signed the waiver form permitting the Census Bureau to request their scores. Three-fourths of those for whom scores were not sought failed to meet the first criterion.

same test by the same publisher are parallel. There is far less consensus regarding alternate tests by the same publisher, and even less regarding alternate tests from different publishers. If tests are parallel, the problem of equating scores is analogous to that of converting centimeters to inches or pounds to grams, i.e., a direct linear transformation of scale. If tests are non-parallel, the problem of conversion is viewed to be more analogous to a conversion from inches to pounds, i.e., a far more complex process involving controversial assumptions about the bivariate distribution of the two variables within particular populations.

In considering the conversion of scores from non-parallel tests Angoff has identified three important questions which must be considered by the investigator:[4]

1. How similar are the tests for which comparable scores are to be developed?
2. How appropriate is the group on whom the table of comparable scores is based when one considers the person or the group for whom the table is to be used?
3. How much error can we safely tolerate in the particular use we have in mind?

Before designing our approach we considered each question carefully. The tests which produced the available scores were all tests of mental functioning, although they were identified by their publishers as tests of "mental ability," "intelligence," "mental maturity," "educational ability," etc. Since such tests as these are often used interchangeably by educators for guidance, selection, and placement purposes we assumed them to be "similar" in Angoff's terms.

---

[4] William H. Angoff, "Can Useful General-Purpose Equivalency Tables be Prepared for Different College Admission Tests?" Proceedings: 1962

The problem of developing a table of comparable scores for different tests requires a procedure which takes into account not only their differing means and standard deviations but also their differing reliabilities and inter-correlations. To develop a meaningful table of comparable scores, data for the same subjects on all pairs of tests for a series of relevant subpopulations (stratified on such important variables as age, sex, and race) are required. Lacking such data we had to make the assumption that the many tests were equally reliable and perfectly correlated and directed our attention solely to the matter of correcting for different means and standard deviations. As is noted below, in spite of its "erroneous" nature this assumption did not prove particularly troublesome.

The issue of tolerable error clearly is different in the case of large-scale social research than in the typical psychometric case. In the typical case the purpose of the conversion is to enable a practitioner (e.g., a college admission officer) to make a decision regarding an individual case (e.g., whether or not an applicant should be admitted to a particular college). In such cases the tolerance for error is necessarily quite small, for the consequences of error for the applicant (although not necessarily for the college) can be rather severe. In social research the investigator typically is interested in the estimation of measures of central tendency for groups or in assessing analytic relationships among variables, and in general such estimates would be far less affected by errors in the conversion process than would individual scores. Therefore, although we assumed the proposed conversion

---

Invitational Conference on Testing Problems, ed. by Eric F. Gardner (Princeton: Educational Testing Service, 1963), pp. 57-73.

procedures to contain tolerable error, we designed an analysis to assess the reasonableness of this assumption.

Conversion Procedures

The scores reported by the educational officials were in a variety of forms. In some instances they were traditional IQ scores, in other cases standard scores, and in still other cases they were reported as percentile scores, percentile bands or stanines. In order to transform all scores to a common metric, information was solicited from the various test publishers regarding the means and standard deviations of the tests reported to the Census Bureau.

As reported by the test publishers the largest number of available scores were based upon a distribution with a mean of 100 and a standard deviation of 16. Therefore, it was decided to use that scale as the common metric. Accordingly, all scores reported in standard score form based upon a different metric were converted to z-score equivalents and then to the common metric. Scores reported in percentile form were also converted to z-scores and then to the common metric. If percentile bands were reported they were "centered" and then converted as in the case of the percentile. Scores reported as stanines were also centered and converted directly.

In order to consider the utility of estimating mental ability from grade point averages in 190 instances where no mental ability score of any type was reported but a grade-point-average (GPA) was, a rough correspondence for that school between mental ability and GPA was estimated from the available data, and a mental ability score on the common scale was computed. In all cases the name of the test, and the method of conversion was noted for later consideration.

Assessment of Comparability

To assess the comparability of the scores derived from the various tests and equated using the procedures described above a series of analyses were carried out similar to those intended for the larger study in which the scores were to be used. Three variables known from previous research to predict mental ability were selected as predictors: father's occupation, father's education, and mother's education. Each of these socioeconomic measures had been developed from responses obtained earlier in the data collection process by the Census Bureau through a standardized interview with each male in the study sample. Thus, they could be considered highly standardized across individuals.

Of the 3,375 individuals for whom test scores were available, only 2,429 were used in the analysis discussed below. The other 946 cases could not be used because information on one or more of the predictors was lacking. The data which are available suggest that relative to the included group, the excluded group somewhat overrepresents youth from disadvantaged socioeconomic backgrounds. For example, the mean number of years of schooling completed by the fathers of those in the excluded group is 9.6, as compared to 10.6 for those included.[5] Thus, it is not surprising that the mean mental ability score of the excluded group is lower than that of the included group, i.e., 96.7 versus 103.4. The consistent direction of these differences supports our belief that excluding the 946 cases did not produce any important distortion in our results.

---

[5]The mean years of father's education for the excluded group is based on 358 cases.

In designing the analysis, seven data groups were constructed from among the individuals whose mental ability scores were available through the conversion process. A score from the Otis Quick Scoring Mental Ability Test was reported for approximately 25 percent of the data cases and so those 635 subjects with scores on that test were treated as a single test group. Similarly, the 443 subjects whose scores were based upon the California Test of Mental Maturity were treated as a distinct group. Since both the Lorge-Thorndike Intelligence Test and the Henmon-Nelson Test of Mental Maturity are administered by the same publisher, and since the number of subjects was relatively small in each case, these two tests were pooled into a single test group for purposes of the analysis. Subjects with scores originally from the PSAT, SAT, and SCAT tests published by the Educational Testing Service were also pooled for similar reasons.

No single test or test publisher was common to more than 20 percent of the remaining 601 subjects and so further, but less precise, pooling was conducted to obtain a fifth and sixth data group. In addition, the scores estimated from GPAs were retained as a seventh data group in order that they be treated separately. The number of cases within each of the seven data groups ranged from 635 in the case of those subjects whose scores were based upon the Otis test to the 190 cases whose scores had been estimated from the reported GPA.

Table A-1 presents the means and standard deviations for the mental ability scores and the three predictor variables within each of the seven test groups as well as within the total sample. There it can be noted in particular that different test groups have somewhat different means on the common measure of mental ability. However, given the

TABLE A-1

MEANS AND STANDARD DEVIATIONS BY TEST GROUP: VARIABLES USED IN ASSESSING THE PROCEDURE OF POOLING MENTAL ABILITY SCORES FROM DIFFERENT TESTS

Test group[a]	N	X_1[b]		X_2[c]		X_3[d]		X_4[e]	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	635	36.7	23.0	10.7	3.3	11.0	2.7	104.8	14.2
2	443	34.4	23.1	10.3	3.7	10.7	2.9	102.4	14.6
3	271	36.6	22.2	10.8	3.1	11.2	2.6	104.2	13.8
4	289	40.7	22.8	11.5	3.2	11.7	2.6	108.1	14.9
5	379	34.9	22.8	10.2	3.4	10.8	2.9	102.6	16.3
6	190	34.9	24.2	9.9	3.9	10.0	3.2	96.1	14.3
7	222	36.5	23.7	10.6	3.5	11.3	2.8	102.0	16.7
Total	2,429	36.3	23.1	10.6	3.5	11.0	2.8	103.4	15.1

[a]Group 1: Otis Quick Scoring Test of Mental Ability
Group 2: California Test of Mental Maturity
Group 3: Lorge-Thorndike Intelligence Test
Henmon-Nelson Test of Mental Ability
Group 4: Preliminary Scholastic Aptitude Test
Scholastic Aptitude Test
School and College Ability Test
Group 5: Miscellaneous additional tests
Group 6: Ability scores estimated from GPA reports
Group 7: Test of Educational Ability
Primary Mental Ability Test
Iowa Test of Educational Development
Differential Aptitude Test
American College Testing Program
National Merit Scholarship Qualifying Test

[b]Father's occupation when the respondent was 14 years of age, measured in terms of the Duncan index of occupational status.

[c]Number of years of formal schooling completed by respondent's father.

[d]Number of years of formal schooling completed by respondent's mother.

[e]Mental ability score.

similar variation among the groups in the means for the three predictor variables these differences in mean measured mental ability would seem to suggest variation in the socioeconomic characteristics of the sub-populations who are administered the various tests, rather than errors in the conversion process. Presented in Table A-2 are the zero-order correlations between each of the three predictor variables and the mental ability scores and between all pairs of the predictor variables. In the case of father's occupation as a predictor of mental ability (r_{14}), the coefficients vary between .24 and .36, but such a range is certainly within the limits of that between father's occupation and the other two predictors where all measures are standard across the seven test groups. Similarly the coefficients for father's education and mental ability vary between .30 and .42 and those for mother's education and mental ability between .22 and .40, but again their range does not seem excessive in comparison with that noted between pairs of the three predictors.

In order to conduct a systematic test of the variations between the different test groups, a series of multiple regression analyses were conducted. Table A-3 presents the resulting coefficients and their levels of statistical significance. Table A-4 contains statistics to test the significance of all paired differences between the same coefficients based upon analyses within different test groups. The statistics in Table A-4 were derived from a series of regression equations in which the regressors included dummy variables for the relevant strata (tests) and products of each of those dummy variables with the continuous predictor variables. Thus, for example, an equation to test for differences

TABLE A-2

ZERO-ORDER CORRELATION COEFFICIENTS, BY TEST GROUP: VARIABLES USED IN ASSESSING THE PROCEDURE OF POOLING MENTAL ABILITY SCORES FROM DIFFERENT TESTS

Coefficient[b]	Test Group[a]							Total sample
	1	2	3	4	5	6	7	
r_{14}	.30	.34	.32	.36	.29	.32	.24	.31
r_{24}	.32	.39	.35	.32	.42	.30	.34	.36
r_{34}	.27	.36	.36	.27	.40	.22	.27	.33
r_{12}	.57	.53	.57	.50	.55	.58	.55	.55
r_{13}	.41	.40	.37	.37	.40	.34	.43	.40
r_{23}	.56	.58	.50	.64	.63	.60	.62	.59

[a] See note a, Table A-1.

[b] X_1 = Father's occupation.

X_2 = Father's education.

X_3 = Mother's education.

X_4 = Mental ability.

TABLE A-3

COEFFICIENTS FOR THIRD-ORDER REGRESSION OF MENTAL ABILITY ON FATHER'S OCCUPATION, FATHER'S EDUCATION, AND MOTHER'S EDUCATION, BY TEST GROUP

Test group[a]	Number of cases	Intercept	Father's occupation	Father's education	Mother's education	R^2 (adj.)	F-ratio
1	635	86.92*	+.098*	+ .703*	+ .618*	.13	32.27*
2	443	81.90*	+.105*	+ .760*	+ .909*	.19	36.02*
3	271	80.26*	+.092*	+ .625*	+1.238*	.17	19.96*
4	289	88.48*	+.173*	+ .618	+ .476	.15	18.02*
5	379	75.46*	+.049*	+1.176*	+1.247*	.21	33.53*
6	190	83.75*	+.125*	+ .523	+ .286	.11	8.76*
7	222	81.96*	+.047	+1.183*	+ .516	.11	10.44*
Total sample	2,429	81.80*	+.097*	+ .818*	+ .863*	.19	21.42*

[a]See Table A-1 for identification of test groups.

*Significant at .05 level or below.

TABLE A-4

T-RATIOS FOR INTER-GROUP COMPARISON OF THIRD-ORDER REGRESSIONS OF MENTAL ABILITY ON FATHER'S OCCUPATION, FATHER'S EDUCATION, AND MOTHER'S EDUCATION, BY TEST GROUP

	Test Group[a]					
	2	3	4	5	6	7
Test Group 1[a]						
Intercept	1.66	1.46	- .35	3.11*	.77	1.09
Father's Occupation	-.14	.12	-1.48	1.04	- .46	.94
Father's Education	-.17	.19	.20	-1.28	.42	-1.13
Mother's Education	-.79	-1.41	.30	-1.58	.72	.21
Test Group 2						
Intercept		.18	-1.61	1.49	- .63	- .19
Father's Occupation		.23	-1.28	1.10	- .34	1.01
Father's Education		.32	.33	1.10	.54	- .98
Mother's Education		-.72	.88	- .81	1.31	.78
Test Group 3						
Intercept			-1.51	1.00	- .68	- .31
Father's Occupation			1.31	.72	- .49	.69
Father's Education			.02	-1.20	.20	-1.11
Mother's Education			1.39	- .02	1.79	1.29
Test Group 4						
Intercept				2.76*	.93	1.20
Father's Occupation				2.24*	.73	2.03*
Father's Education				-1.21	.19	-1.18
Mother's Education				-1.49	.33	- .07
Test Group 5						
Intercept					-1.90	-1.36
Father's Occupation					-1.22	.05
Father's Education					1.38	.02
Mother's Education					1.93	1.39
Test Group 6						
Intercept						.35
Father's Occupation						1.14
Father's Education						-1.28
Mother's Education						- .40

[a] See Table A-1 for identification of test groups.

* Significant at .05 level.

between the coefficients in stratum 1 and stratum 2 would take the following form:

$$(A.1)\ X_4 = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4D + a_5(DX_1) + a_6(DX_2) + a_7(DX_3) + u,$$

where D = 1 for observations in stratum 2, and

= 0 otherwise.

The six regressions of this type which were performed were much more elaborate because they were designed to test for differences among all of the strata. Thus, the general form of Equation (A.1) was as follows:

$$(A.2)\ X_4 = a_{k0} + a_{k1}X_1 + a_{k2}X_2 + a_{k3}X_3 + \sum_{i\neq k}^{7} a_iD_i + \sum_{i\neq k}^{7}\sum_{j=1}^{3} a_{ij}(D_iX_{ij}) + u,$$

where the kth stratum is the reference stratum to whose coefficients the coefficients of the other six strata were compared.[6]

In general there is very little evidence which suggests that the intercepts or regression coefficients resulting from the analyses within the different test groups are from different populations. Of the 84 coefficients presented in Table A-4 (21 pairs of test group comparisons for 4 parameters) only four are statistically significant at the .05 level, and in no case does the comparison between any two data groups produce more than two coefficients whose difference is statistically significant. Further, two of the four significant differences are with

---

[6] See also Damodar Gujarati, "Use of Dummy Variables in Testing for Equality between Sets of Coefficients in Linear Regressions: A Generalization." The American Statistician, XXIV (December 1970), pp. 18-22.

respect to the intercept, which given the somewhat different group means on the three predictors suggests more a difference among the subpopulations administered the different tests than among the internal properties of the test scores themselves.

The comparison between Group One (Otis test) and Group Two (California test) is particularly interesting for these groups represent the two most frequently reported tests for this national sample. Neither the intercepts for these two tests nor any of the regression coefficients differ significantly. In considering the comparison between Group Three (Houghton-Mifflin tests) with Group Four (Educational Testing Service tests), the same negative findings can be observed. The case in which two significantly different coefficients occurs is that between Group Four (Educational Testing Service tests) and Group Five (a potpourri of little known and often only locally used tests). Given the rather different nature of these two test groups on the three socioeconomic indicators (Table A-1), it does not seem unreasonable that even on a common test of mental ability the Group Four intercept would be in excess of that for Group Five, or that the regression coefficient for father's occupation would also vary.

The four statistically significant pair-wise differences are depicted graphically in Figures A-1 and A-2. Each curve on the graph represents a normal density (frequency) function defined by the value of a regression coefficient and its standard error. In addition, each curve is traced out with the numeral of the stratum to which it applies. The amount of common area under any two curves indicates the level of confidence which we have in accepting the hypothesis that the effect of a variable on mental ability is the same for both strata. Thus, Figure

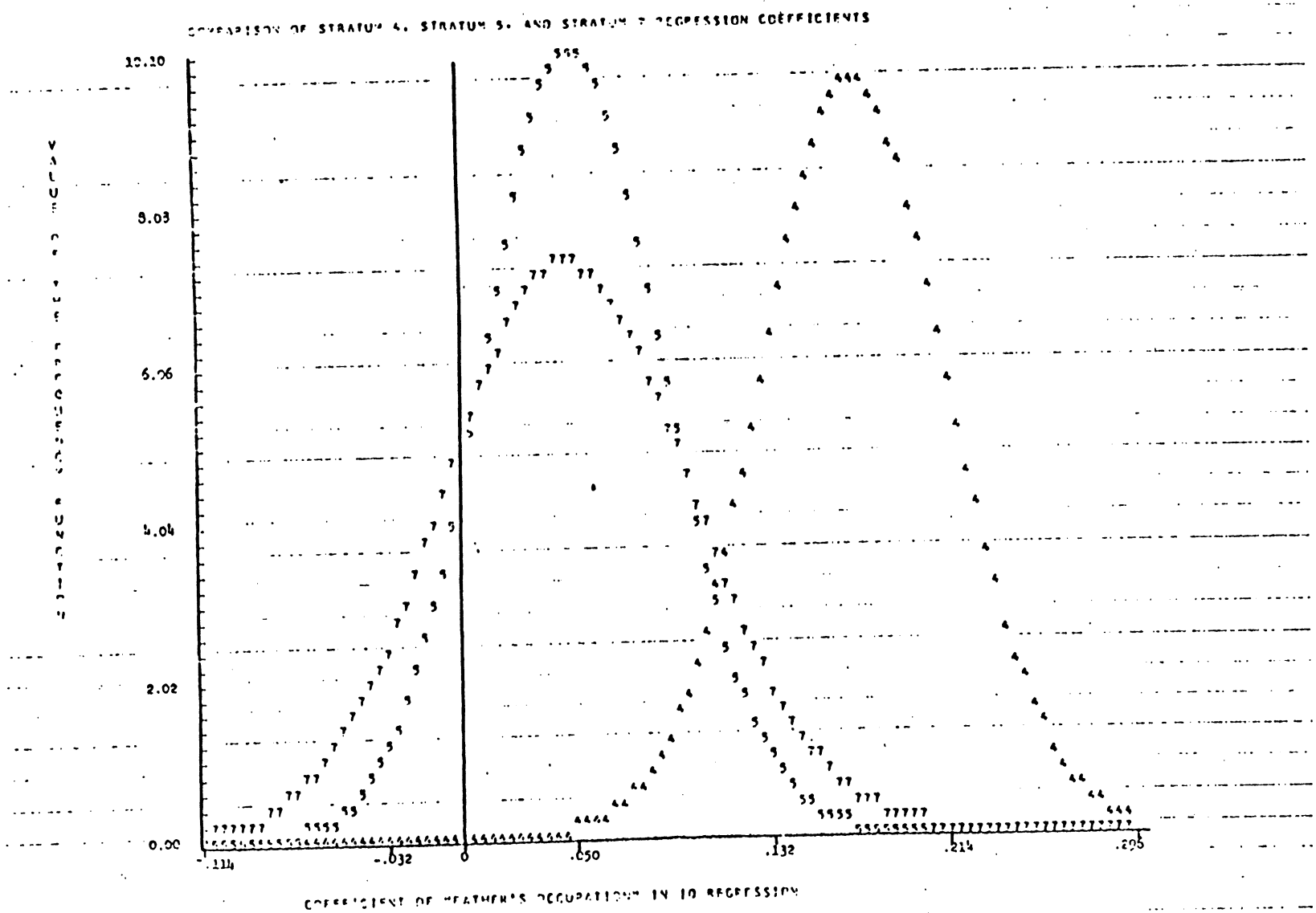

Figure A1

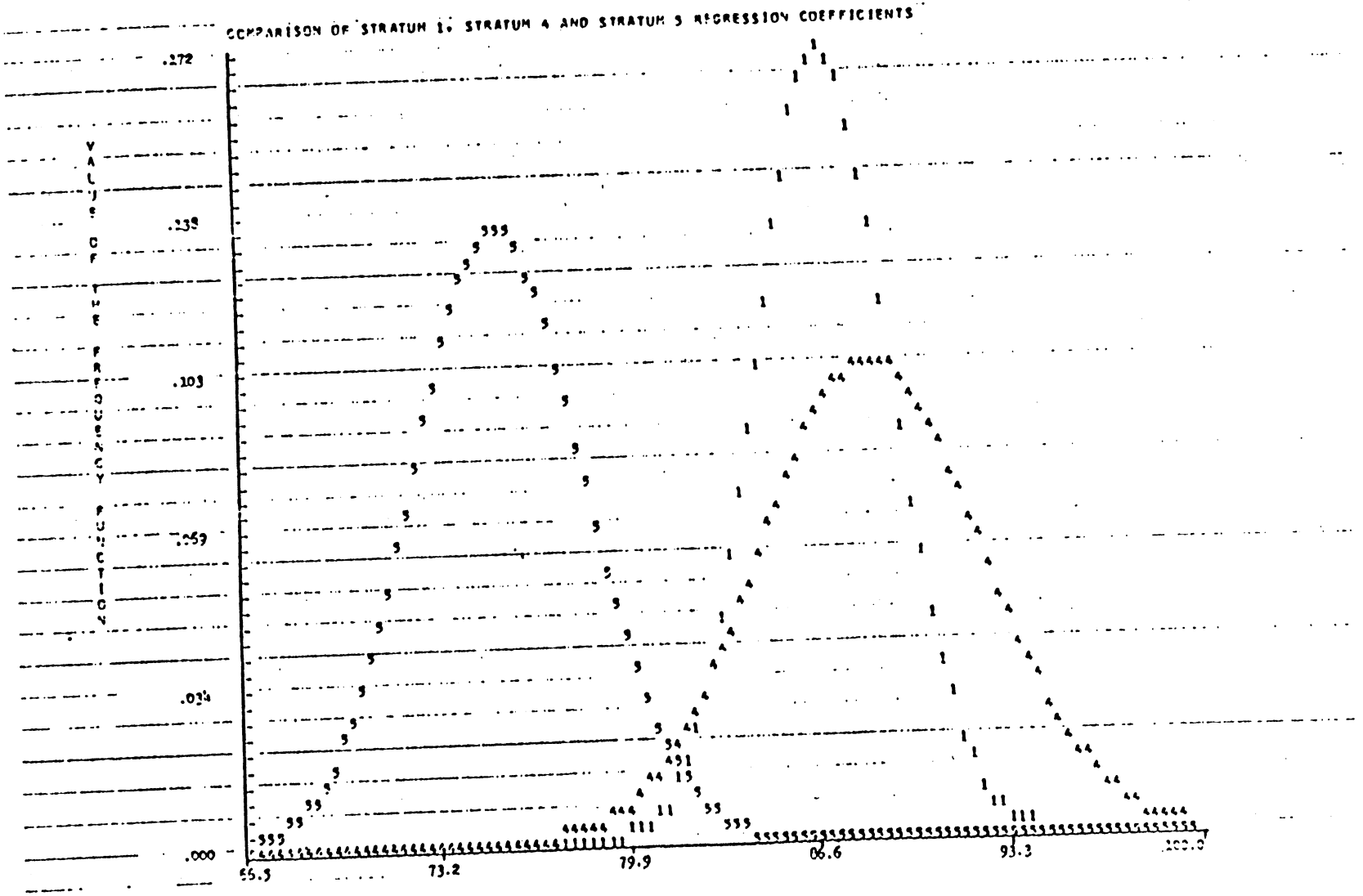
COMPARISON OF STRATUM 1, STRATUM 4 AND STRATUM 5 REGRESSION COEFFICIENTS
.272
.239
.103
.069
.034
.000
VALUE OF THE FREQUENCY FUNCTION
55.5
73.2
79.9
86.6
93.3
100.0
INTERCEPT OF IU REGRESSION

Figure A2

A-1 clearly indicates why the hypothesis of equality of the effect of father's occupation is rejected in the comparison between stratum 4 and stratum 5, and in that between 4 and 7. The large common area under the curves for strata 5 and 7 similarly indicates why the hypothesis is accepted in this instance.[7] Analogous inferences can be drawn from the pair-wise comparisons of intercepts among strata 1, 4, and 5 as illustrated in Figure A-2.

Allowing for differences among the data groups in sample size, it would seem that regardless of the particular mental ability test from which a common score was obtained, the relationship (as measured by the regression coefficients) between socioeconomic status and mental ability is remarkably similar. Although certain coefficients are significant within some data groups and not within others, the joint effect of the three predictor variables as measured by an R^2 adjusted for degrees of freedom is significant in all cases (Table A-3). Particularly important is the fact that both the zero-order and third-order coefficients for the pooled sample appear well within the limits reported in past research using a single test of mental ability.

Implications

On the basis of these results we see little reason for social scientists engaged in analytic research on national samples of youths or young adults to be reluctant to pool data from different commonly used

---

[7] The location of the zero point on the horizontal axis in this graph indicates why the coefficients of father's occupation in strata 5 and 7 were judged to be insignificantly different from zero (see Table A-3). It is clear that the area to the right of the zero line comprises much less than 95 percent of the area under the curve in each case.

tests of mental ability after first correcting for their varying means and standard deviations. Certainly the error introduced by such a procedure for "equating" non-parallel tests seems small in comparison with the value of having a measure of mental ability available for analysis.

However, in addition to the procedures utilized in the present study it seems desirable to attempt to make a greater provision in the equating process than was possible in the present study for possible varying inter-correlations between pairs of the different tests. Therefore, we suggest that instead of asking for data on only the most recent test, as was done in the study upon which our analysis was based, future investigators obtain data on as many of the seven most frequently used tests of mental ability as are available within a school's records. Not only will this minimize the number of different tests whose scores must be transformed to a common metric, but it will also permit the estimation of inter-test correlations which can be introduced as weights into the conversion process.

In some instances the school may not have available a score from one of the seven tests, but may have a score from some other test. To maximize response, it seems advisable to ask for such a score as well. However, since such a score will have to be handled with special care, at the time of data processing a decision will have to be made regarding whether or not, given the frequency of such occurrences, the objectives of the survey warrant the additional cost of manually coding and transforming such scores to the common metric.

A suggested format which accomplishes these objectives is presented in Figure A-3. In addition, in order to assure the release of test information by school officials it is recommended that written permission

FIGURE A-3

PROPOSED FORMAT FOR OBTAINING INDIVIDUAL MENTAL ABILITY SCORES FROM SCHOOL RECORDS

_____
(Name of Individual)

Do you have a record of any group administered SCHOLASTIC APTITUDE or INTELLIGENCE test score OR national percentile for this person?

1 ___ Yes -- continue with questions 1a & 1b

X ___ No -- skip to question 2

a. For EACH of the following tests please record the most recent test scores and national percentiles for this person. (If for any tests either the score or the percentile is unknown please write "NA" (i.e., not available) in the appropriate space.)

	Name of Test	IQ Score	National Percentile
(01)	California Test of Mental Maturity (CTMM)	____	____%ile
(02)	Otis Quick Scoring Mental Ability Test	____	____%ile
(03)	Lorge-Thorndike Intelligence Test	____	____%ile
(04)	Henmon-Nelson Test of Mental Maturity	____	____%ile
(05)	Kuhlmann-Anderson Intelligence Test	____	____%ile
(06)	Differential Aptitude Test (DAT)	____	____%ile
(07)	School and College Ability Test (SCAT)	XXXXX	____%ile

b. If this person has not taken any of the above seven tests but has taken some other aptitude or intelligence test, please give the name of the most recent test and the appropriate scores.

Name of Test	IQ Score	National Percentile
_____	____	____%ile

---

for access to these data be obtained from the subject prior to the time of inquiry and forwarded to the school at the time the request for test data is made.

It is our estimate that with a format of the type proposed in Figure A-3 and with statements for the release of the data, machine transformable test scores can be obtained for at least 90 percent of a national sample of subjects still enrolled in school. For subjects not enrolled in school the percentage would of course be less, but in the urban areas school officials seem to be able to retrieve test data for persons up to 24 years of age. With the increased use of automated storage and retrieval systems by other school systems, increasingly such data should be accessible for additional subjects.