DOCUMENT RESUME

ED 065 514 TM 001 434

AUTHOR	Hilton, Thomas L.
TITLE	Predictability and Intellectual Growth--Some Comments on the Degree and Interpretation of Growth Correlations.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	RM-72-1
PUB DATE	Feb 72
NOTE	14p.; Expanded version of paper presented at annual meeting of American Psychological Association, Washington, D.C., 1971
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Achievement; *Achievement Tests; *Correlation; Elementary School Students; *Grade Equivalent Scores; Predictive Ability (Testing); Secondary School Students; *Test Interpretation; *Test Results
IDENTIFIERS	SCAT; School and College and Ability Test; Sequential Test of Educational Progress; STEP

ABSTRACT

Five possible interpretations are given of very high correlations between scores on successively administered ability tests in a longitudinal sample of approximately 7,000 public school students tested in grades 5, 7, 9, and 11. At each of the four grades, students were given the appropriate level of the Sequential Test of Education Progress (STEP) and the School and College and Ability Test (SCAT). The correlation between grade levels of a verbal factor and a quantitative factor were: verbal factor, .94 (5th grade vs. 7th), .95 (7th vs. 9th), and .96 (9th vs. 11th); and for the quantitative factor, .90 to .93 to .95. The interpretations are: (1) During these two-year periods, U.S. students change intellectually very little; (2) The high correlations result from methods or from factors specific to each SCAT and STEP test; (3) The high correlations result from the tests' measuring general intellectual abilities which mature without being influenced by differential student experience; (4) Which school a student attends makes no difference; and (5) Each student's growth rate is set early in his life and remains constant thereafter. None of the five interpretations were found to be wholly acceptable. It is concluded that suitable measures of all variables related to the data analysis of each of the probable causal pathways involved in the growth process in question are needed. (DB)

ERIC

ED 065514

RM-72-1

RESEARCH MEMORANDUM

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS OOCUMENT HAS BEEN REPRO-OUCEO EXACTLY AS RECEIVEO FROM THE PERSON OR ORGANIZATION ORIG-INATING IT. POINTS OF VIEW OR OPIN-IONS STATEO OO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EOU-CATION POSITION OR POLICY.

PREDICTABILITY AND INTELLECTUAL GROWTH--SOME COMMENTS ON THE DEGREE AND INTERPRETATION OF GROWTH CORRELATIONS

Thomas L. Hilton

This is an expanded version of a paper presented at the 1971 Annual Meeting of the American Psychological Association, Washington, D. C.

TM 001 434

Educational Testing Service
Princeton, New Jersey
February 1972

Predictability and Intellectual Growth--Some Comments on the Degree and Interpretation of Growth Correlations[1]

Thomas L. Hilton

Assume for the moment that some elementary school test scores are very highly correlated with high school test scores obtained from the same students six years later. By "very high" we mean correlations in the vicinity of .90. How does one interpret such a finding? Does it mean, for example, that a student's elementary school achievements are the major determinants of his subsequent academic growth? Are family and school variables relatively less important? Does it mean that formal schooling "doesn't make a difference" or that the particular school which a student attends doesn't make a difference? Finally, does it mean that students do not change from the fifth to the eleventh grade? These questions are the subject of this paper.

The predictability in question was examined in the Growth Study, a nation-wide study of academic growth undertaken by Educational Testing Service in 1961 (Anderson & Maier, 1963; Hilton & Myers, 1967). As part of that study, achievement test scores were obtained for a longitudinal sample of approximately 7,000 public school students tested in grades 5, 7, 9, and 11. At each of these four grades the students were given the appropriate level of the Sequential Test of Educational Progress (STEP) and the School and College and Ability Test (SCAT). The correlation between the grade 5 scores and the grade 11 scores can be described in a number of different ways.

---

[1]The author is indebted to Charles E. Werts for helpful criticism of an earlier draft of this paper. This paper is an expanded version of a paper presented at the 1971 annual meeting of the American Psychological Association, Washington, D. C.

The first description of the correlation is provided by Table 1 which shows the joint distribution of grade 5 and grade 11 composite scores by quintiles. The composite score is an unweighted sum of each student's two SCAT scores and six STEP scores. The table shows that, of the 790 students in the lowest quintile in grade 5, 70% were still in the lowest quintile in grade 11, 21% had moved up to the second quintile, 7% to the third quintile, 2% to the fourth, and 0% to the top quintile. (Actually there was one student out of 472 who moved from the lowest quintile to the highest.) The students in the highest grade 5 quintile were even more stable. Seventy-six percent remained in the top quintile; 20% dropped to the fourth quintile, 4% to the third, 1% to the second, and 0% to the bottom quintiles. (The frequency was five in the second quintile and 0 in the lowest quintile.)

The general picture in Table 1 is one of high correlation between the grade 5 and grade 11 scores. The product-moment correlation between the grade 5 composite scores and the grade 11 composite scores is .85 for the total sample. This correlation is, incidentally, slightly higher for the girls alone (.87) than for the boys alone (.84), even though the standard deviations of the grade 5 and grade 11 distributions were slightly higher for the boys (9.1 and 8.9) than for the girls (8.5 and 8.4). For the white students alone the correlation is .83 and for the black students, .79. The correlation for the total sample is larger than these, presumably because the pooled distributions have a larger standard deviation than either racial sample alone.

These correlations are high, but still are underestimates of the true correlations, i.e., correlations between error free measures. Jöreskog (1969), using his general model for the analysis of covariance structures,

Table 1

Grade 11 Standing of Students on Composite Score Scale

Grouped by Grade 5 Standing[a]

		% of Grade 5 Students in Each Grade 11 Quintile				
Grade 5 Quintile	N	Lowest	2nd	3rd	4th	Highest
Lowest	790	70	21	7	2	0
2nd	798	23	43	23	8	2
3rd	797	6	27	38	25	4
4th	789	0	8	27	45	20
Highest	792	0	1	4	20	76
Total	3966					

[a] r = .8530.

factored these same data and decided upon two familiar factors at each grade level, a verbal factor and a quantitative factor. The correlation between grade levels of these factors gives us an estimate of the true correlations underlying these data. For the verbal factor the correlations increased from .94 (5th grade vs. 7th grade) to .95 (7th vs. 9th) to .96 (9th vs. 11th), and for the quantitative factor the corresponding correlations increased from .90 to .93 to .95. By anyone's measure these correlations are very high, leaving precious little variance in any set of scores which is not explained by an earlier set of scores. How then do we interpret them. A number of possible interpretations will be discussed. Some of the interpretations are admittedly straw men which would be omitted were it not that examples of such misinterpretations can be found in the research literature.

Interpretation 1. During these two-year periods, U. S. students change intellectually very little. There is, of course, no basis for concluding this from the correlations reported. As every beginning student of statistics learns, correlations tell nothing about changes in variation or mean gain. A perfect correlation would be consistent with a drastic increase in the differences among students and/or with considerable gain by the group as a whole.

In actuality the mean SCAT and STEP scores do increase. From the 9th to the 11th grade, for example, this increase in the converted scores averages about seven points on each test (Hilton & Patrick, 1970). This is approximately one-half the standard deviation of the 9th grade scores. Thus the average 11th grader achieves a higher score than approximately 70% of the 9th graders. In terms of the items on individual tests the 11th graders successfully answered about five more items than the 9th graders, the raw score to scale score conversion being roughly 1 to 1 1/2.

Is this gain a lot or a little? By what percentage do students increase their knowledge from the 9th to the 11th grade? Unfortunately this question is unanswerable without ratio scales of ability. It is clear, however, from examining the content of the tests at different levels along with normative data that the typical student does grow intellectually from the 5th to the 11th grade.[2] What does remain relatively invariant is the relative standing of the students from one grade level to the next. Correlations of the magnitude reported above leave little room for changes in the ordering of the students.

Interpretation 2. The high correlations result from methods or form factors specific to each SCAT and STEP test. The methods factor (Campbell & Fiske, 1959) might result, for example, from consistencies from one level to the next level in the format of the tests. According to this interpretation, the high correlations result from the similar way in which the tests at different levels are assembled and administered.

Actually, the analysis by Jöreskog (1970) partially anticipated this possibility. It allowed for--and obtained--methods factors specific to each test and these factors were independent of the grade level factors mentioned above. However, any methods variance which was common to all the SCAT and STEP tests would appear in the grade level factors. Thus the factor correlations reported above may reflect some of a methods factor and Interpretation 2 cannot be rejected although intuitively it seems unlikely that the high correlations could be explained entirely on these grounds.

---

[2]Shaycoft (1967), in a longitudinal study at the high school level, also concluded that students grow. She found that the gains "are uniformly in the right direction...and in the more important areas they are quite substantial in magnitude."

Interpretation 3. The high correlations result from the tests' measuring general intellectual abilities which mature without being influenced by differential student experience. The rank ordering of the students remains relatively stable even though the abilities develop from grade 5 to grade 11.

This interpretation also cannot be rejected. The STEP and SCAT tests were broadly conceived. The STEP tests were designed by a panel of teachers to measure skills and understanding of basic importance in education. The emphasis in the tests is on applying knowledge and skills to new situations rather than on memory for facts. In order that the tests be widely useful in a broad range of schools they emphasize general, widely taught principles. Thus the composite scores and the factor scores mentioned above were derived from tests which are highly similar in conception. What is measured is more like what is commonly referred to as ability than achievement, for which reason that term is used in this paper. If the tests were more oriented to specific learning outcomes one might see more changes in relative position.

A second aspect of the instruments is also relevant. Tests of this type can be designed to measure the cumulative knowledge and skill of the students as it has developed over the years or they can be designed to focus on items reflecting those skills which are most likely to have changed since earlier administrations of lower forms of the test. In the latter case the item selection method is designed to select so-called "change items" (Bereiter, 1962). The SCAT and STEP items were not selected in this way. Each level of the test includes the knowledge and skill measured in lower levels of the test. Thus there is to some extent a built-in correlation between scores from successive test administrations.

An important question is whether this measurement model (i.e., SCAT and STEP) and this statistical model (product-moment correlations) accurately simulate student academic growth. As mentioned above, the tests were designed by teams of teachers who, to the best of their knowledge, defined what the students at each level should know and be able to do. If the content of successive tests overlap, then this to some extent reflects the way the world is. As for the statistical model, we are--in computing correlations between scores at two points in time--assuming that the general linear model is an appropriate way to describe the relationship between ability at one time and ability at a later time.

Lastly, if we had instruments measuring educational outcomes other than academic ability, e.g., changes in self-perception, in individual goals, values, and attitudes, then again we might observe more differential change. But these are suppositions. For the time being we cannot reject the available evidence which indicates that the true rank ordering among students in academic ability changes very little in two-year periods and only slightly more so in a six-year period and that this stability could be attributable to the design of the instruments.

Interpretation 4. Which school a student attends makes no difference. The argument here would be that the 5th, 7th, 9th, and 11th grade test scores are so highly correlated (or at least the factor scores are) that the proportion of variance possibly attributable to the school must be very small. There is an alternate possibility, however. This is that the scores from successive grades are both influenced by a third variable--a school characteristic, for example--and thus that the high correlation in question is partly spurious as far as any direct relationship between successive scores is concerned.

The same argument applies to possible effects of educational innovation. If receiving a special treatment or not is correlated with both initial and final scores (perhaps only affluent students who tend to have high initial scores and high final scores receive the treatment), then a high initial-final correlation again will be in part spurious.

Still another possibility arises when the special-treatment group is small in number relative to the rest of the sample. Perhaps one school in a sample of 25 received the treatment. The variance contributed by the treatment is, then, unlikely to change the initial-final correlation appreciably.

Werts and Linn (1970) have examined the implications of the various statistical models in this area. For our present purposes the important point is that the zero-order correlation between two successive test administrations does not permit us to say whether an external variable, e.g., school attended, or an educational innovation, influenced the growth in question. Inferences of this type require that all major sources of influence be specified and that the analysis encompass all of the relevant variables. Typically we do not know all the relevant variables and, further, do not have adequate measures for many of those we do know. But we should keep in mind that to the extent that relevant variables are omitted our results may be misleading.

Interpretation 5. Each student's growth rate is set early in his life and remains constant thereafter. A number of reasons might be hypothesized for such lack of variation in growth. Our schools may be administered so as to preserve the rank order among students. The better students may consistently receive the better teachers. Tracking systems and homogeneous grouping may contribute to fixity of growth rates. Early labeling of students in accordance with their measured ability may be self-fulfilling. Each of these charges has been made by one or another critic of the schools.

Another possibility is what might be called the Talents Hypothesis, from the Parable of the Talents.[3] The students who, for one reason or another, are more knowledgeable gain more from a given amount of learning effort than their less sophisticated classmates, while the low-ranking students are progressively more handicapped by their partial knowledge. The result is that the initially high scoring students pull even farther ahead on subsequent test administrations.

Finally, the stability may be biological in origin or attributable to early childhood experiences. In any case we again find that the correlation between successive test administrations is not relevant evidence. Let us assume that learning is cumulative and that from the 9th to 11th grade of high school the typical student adds an increment to his cumulated learning which is small relative to that which he learned prior to entering high school. It follows that most of the 11th grade score represents knowledge that he had in the 9th grade. Given this overlap the correlation between ability at grades 9 and 11 will be high even if the gains from grade 9 to grade 11 are <u>random</u> increments, as Anderson (1939) pointed out. Thus we find again that the predictability of test performance from earlier test performance is by itself a theoretically uninterpretable finding.

What is of more interest in this context is the correlation between ability at one grade level and the gain in ability in subsequent grades. But here again

[3] "For unto everyone that hath shall be given, and he shall have abundance but from him that hath not shall be taken away even that which he hath." St. Matthew, Chap. 25.

there is the possibility that any correlation is attributable to the correlation of both measures (i.e., the status measure and the gain measure) with an external variable. To assert that prior knowledge somehow determines later growth the researcher is obligated to demonstrate that any correlation between status and gain is not attributable to school, family, or community variables. From this point of view, Thorndike's (1966) focus on only the correlation between intellectual status and intellectual growth without consideration of any external variables is unduly restrictive. For the reasons cited, one cannot draw any conclusions about whether status influences gain when one only has the correlation of status with gain.

Conclusion

We have considered five possible interpretations of very high correlations between scores on successively administered ability tests. Of the five interpretations none was wholly acceptable. The high correlations do not mean that students do not change; they mean that the students' relative standing on the measures in question remains very nearly the same from one grade level to the next. The stability may be attributable to a methods factor, or it may be attributable to the fact that the tests were designed to measure general problem-solving skills and general principles which may be relatively uninfluenced by a student's school experience. It cannot be said from the results cited that which school a student attends has no effect on his growth; the high correlations could result from a school characteristic having a strong effect on both initial and final achievement. Finally we cannot assert from the available data that growth rates are fixed early in a student's career, either for physiological or environmental reasons. This leaves us able to say only that in the abilities measured in this study the ordering of the students remains remarkably stable from the 5th to the 11th grade. Which of the several possible explanations for the stability is most valid is unknown. Nevertheless, that the ordering of students changes so little is a significant fact which raises important educational questions. Is such stability consistent with the goals of American education?

The more general question raised in the first paragraph of this paper concerned the contribution of high predictability to one's understanding of student growth. The particular correlations and the alternative interpretations which were considered suggest that initial-final correlations or status-gain correlations--no matter how high--shed little light on the determinants of growth. What is required are suitable measures of all the variables which are likely to be related to the growth of interest and consideration in the data analysis of each of the probable causal pathways involved in the growth process in question.

References

Anderson, J. E. The limitations of infant and preschool tests in the measurement of intelligence. Journal of Psychology, 1939, 8, 351-379.

Anderson, S. B., & Maier, M. H. 34,000 pupils and how they grew. Journal of Teacher Education, 1963, 14, 212-216.

Bereiter, C. Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1962. Pp. 3-21.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Hilton, T. L., & Myers, A. E. Personal background, experience and school achievement: An investigation of the contribution of questionnaire data to academic prediction. Journal of Educational Measurement, 1967, 4(2), 69-80.

Hilton, T. L., & Patrick, C. Cross-sectional versus longitudinal data: An empirical comparison of mean differences in academic growth. Journal of Educational Measurement, 1970, 7(1), 15-24.

Jöreskog, K. G. Factoring the multitest-multioccasion correlation matrix. Research Bulletin 69-62. Princeton, N. J.: Educational Testing Service, 1969.

Jöreskog, K. G. A general method for analysis of covariance structures. Biometrica, 1970, 57(2), 239-251.

Shaycoft, M. E. The high school years: Growth in cognitive skills. Pittsburgh, Pa.: American Institutes for Research and School of Education, University of Pittsburgh, 1967.

Thorndike, R. L. Intellectual status and intellectual growth. Journal of Educational Psychology, 1966, 57(3), 121-127.

Werts, C. E., & Linn, R. L. A general linear model for studying growth. Psychological Bulletin, 1970, 73(1), 17-22.

ERIC