The Study of Mathematically Precocious Youth

DANIEL P. KEATING, PH. D.[1]
University of Minnesota

One of the enduringly refreshing aspects of the history of psychology, particularly the British and the American disciplines, has been its concern with and attention to individual differences. This interest was accurately expressed by Sir Francis Galton (1889, p. 62) in a statement about those who would ignore such differences: "Their souls seem as dull to the charm of variety as that of the native of one of our flat English counties, whose retrospect of Switzerland was that, if its mountains could be thrown into its lakes, two nuisances would be got rid of at once."

In our Study of Mathematically and Scientifically Precocious Youth at The Johns Hopkins University, individual differences are of monumental importance. Although the group of subjects we have found thus far is narrowly defined by age and precocity in specified areas, there is considerable variability among them. Later I shall discuss this interindividual variability in more detail. Lynn Fox will also be concerned with these differences and their impact on the predictions of success in challenging courses.

In general, however, several things might be pointed to as quite characteristic of the group as a whole. Without becoming enmeshed in the "nature-nurture" controversy, we may fairly say that these students' innate (i.e., genotypic) abilities, both general and specific, are well above average, and that the environment which nourished them provided the necessary interactions for at least partial phenotypic expression. These terms should be interpreted with regard to the "reaction-range concept" (Gottesman, 1963). Genotypes are expressed phenotypically in a specific environment; the range of phenotypic expression for a given genotype may be large. For these students, however, aspects of the total environment, particu-

[1]Now at the Institute of Child Development, University of Minnesota, Minneapolis, Minnesota 55455.

larly the school environment, are usually far from ideal, and part of the purpose of the study is to examine the ways in which they can be made to approximate the ideal more closely.

PRECOCITY AND LATER ACHIEVEMENT

Stanley has examined the concept and the history of precocity in his chapter; thus we need not do so further here. But it should be carefully noted that when much greater than average ability flourishes in an adequate environment, the empirically observable result is often early expression of those abilities through some assessable performance.

This statement implies several caveats which should be spelled out. It does not contradict the possibility of latent talent of the sort which is ascribable to insufficient or inappropriate environmental stimulation or to the confounding effects of personality or other factors which might prevent the manifestation of such talent. Also, precocity in mathematics or science is certainly not a sufficient and possibly not even a necessary condition for eventual achievement in those areas. It is often indicative of great interest in those subjects, however, and it provides a base for the rapid and successful development of achievement in those areas. It is difficult to make a psychological argument from historical sources, but the careful and thorough research of Cox (1926) and Lehman (1953) would seem to indicate that not only is there a history of precocity in the childhood years of a large percentage of generally accepted geniuses but also that this was frequently expressed in some sort of early achievement as well.

The advantage gained by such early precocity may, however, be severely attenuated or lost altogether if it is not properly nurtured. It would seem that lack of "intellectual nourishment" in school is the rule rather than the exception for these special students when active intervention by some agency outside the school (e.g., parents) is absent. Fox (see her article) discusses these issues further, as well as the methods we are learning for nurturing mathematical and scientific precocity.

Thus, we are not neglectful of these conditions of necessity and sufficiency but have attempted to concentrate our efforts on those students who *have* achieved high scores on measures we employ rather than on those who may at some other time, past or future, attain high scores. Some of the reasons for this decision have already been discussed by Stanley (see his article). The result has made the pertinent analyses more complex in some ways and simpler in others. The group is easily defined by reference to specific criteria: for example, Scholastic Aptitude Test Mathematical (SAT-M) score greater than or equal to 600 while still less than 14 years old. Only 11% of high school seniors would score that high.

But we must be aware also for purposes of analysis that we are dealing with one extreme of the distribution of developed talent, at least the upper ½ of 1% of the age group.

FINDING PRECOCIOUS TALENT

The first problem then was to find these precocious students in some systematic fashion. Informal methods such as teacher or parent references proved insufficient, so a mathematics and science competition was organized for seventh and eighth graders. There was no official screening beforehand, but a number of would-be contestants dropped out after receiving the SAT practice booklet and working the practice test. The total number of test takers was 450, with 396 taking math,

TABLE 1
MEAN SCORES OF STUDENTS ON THREE MEASURES: SAT-M, M-I, AND SCIENCE[a]

Tests taken	Test	Statistic	7th Grade Female	7th Grade Male	8th Grade Female	8th Grade Male	9th Grade Female	9th Grade Male	All examinees
		N	59	67	63	67	–	2	258
Math only	SAT-M	Mean	416	450	472	516	–	620	466
		S.d.	71	104	76	100	–	42	97
	M-I	Mean	393	398	431	451	–	540	420
		S.d.	46	63	48	76	–	28	65
		N	18	23	32	62	1	2	138
Math and science	SAT-M	Mean	448	487	427	531	510	760	492
		S.d.	86	100	104	109	–	42	115
	M-I	Mean	407	437	414	467	500	695	445
		S.d.	48	86	52	86	–	106	84
	Science 1A & 1B	Mean	74	83	69	88	100	106	81
		S.d.	13	19	16	21	–	3	20
		N	7	13	4	29	1	–	54
Science only	Science 1A & 1B	Mean	65	66	69	80	103	–	74
		S.d.	7	13	19	18	–	–	17

[a]The total number of students taking the mathematics tests was 396. The total number taking the science test was 192. *Both* sets of tests were taken by 138 students.

192 taking science, and 138 taking both.

In Table 1 are listed the mean scores on SAT-M, Mathematics Achievement Level I (M-I), and STEP II Science (1969) of all the contestants, grouped by sex, grade, and test(s) taken. Perhaps the single most important finding of our study thus far, and one which we are inclined to overlook because we have become acclimated to it, is that there is a remarkable number of almost unbelievably able and academically accomplished young students in grades seven and eight. (Accelerated ninth graders were also eligible.) The level of their ability can be inferred from Table 1, but the picture becomes even clearer when we look at the highest scorers within this able group.

As noted above, a score of 600 on SAT-M places one at the 89th percentile of male high school seniors. In the group of 396 students who participated in the competition 53, or 13%, scored 600 or higher, and 23, or 6%, scored at or above 650, which is the 94th percentile. On M-I, which is an achievement test for high school seniors who have taken seven or more semesters of mathematics, a score of 550 is about the 41st percentile. Twenty-three (6%) of these students scored 550 or greater. The grouped frequency distributions on SAT-M and M-I are given in Table 2. Clearly, whether aptitude *or* achievement tests are used to measure the ability of these students, the best of them are competitive with superior high school seniors. Although it is with these highest scoring students that we will be actively involved, since they are the truly precocious ones in this group, it is important to look at the original group of 450 students more closely.

There is clearly a difference in ability as measured by all three tests between those students who chose to compete in both the mathematics and science contests and those who chose to take only math or only science. The 138 math and science contestants were significantly better than the 258 math only contestants on both SAT-M (0.02

TABLE 2
GROUPED FREQUENCY DISTRIBUTION BY GRADE AND SEX ON SAT-M AND M-I OF 396 STUDENTS PARTICIPATING IN MATHEMATICS CONTEST

Score	7th Grade				8th Grade				9th Grade[a]			
	SAT-M		M-I		SAT-M		M-I		SAT-M		M-I	
	B[b]	G[c]	B	G	B	G	B	G	B	G	B	G
760-800	0	0	0	0	1	0	0	0	1	0	1	0
710-750	2	0	1	0	2	0	2	0	1	0	0	0
660-700	2	0	1	0	13	0	1	0	0	0	0	0
610-650	3	0	0	0	17	0	6	0	1	0	1	0
560-600	8	3	0	0	19	11	8	0	1	0	1	0
510-550	11	8	6	3	21	23	15	9	0	1	1	0
460-500	20	16	9	5	19	20	27	17	0	0	0	1
410-450	14	17	27	22	17	14	33	34	0	0	0	0
360-400	18	22	31	36	13	14	32	29	0	0	0	0
310-350	7	7	11	10	6	7	4	6	0	0	0	0
260-300	3	2	4	1	1	3	1	0	0	0	0	0
210-250	2	2	0	0	0	3	0	0	0	0	0	0
N	90	77	90	77	129	95	129	95	4	1	4	1
Median	457	420	394	388	534	470	442	421	690	510	590	500
Mean	460	423	408	396	523	457	458	426	690	510	618	500
S.d.	104	75	71	49	105	88	81	50	88	–	110	–

[a]Accelerated ninth graders were eligible, i.e., those not yet 14 at time of testing (March 4, 1972).
[b]Boys.
[c]Girls.

$< p < 0.05$) and M-I ($0.001 < p < 0.01$) and also significantly better than the 54 science only contestants on science ($0.01 < p < 0.02$).

A strong "self-concept" factor, which is fairly accurate, may tentatively be inferred to be operating in the self-selection for which test(s) to take. Those students who do in fact score higher than other gifted youngsters on tests such as these would appear to perceive themselves as being knowledgeable in more than one area.

One of the striking and unexpected differences which emerged from this large screening session was the sex difference in mathematical precocity, as can been seen in Tables 1 and 2. There were 43 boys with a SAT-M score of 610 or greater, whereas the three highest scoring girls earned 600. Astin's paper deals with this matter in greater detail.

Parenthetically, the science test was relatively inefficient in screening in the level of talent we are seeking; additionally, quantitative aptitude and achievement seem to be a growing necessity for scientists. Thus, we are dropping the science tests in our future large screening sessions. Those who are screened in on the basis of mathematical talent will, of course, be subsequently tested for their knowledge of general science.

Approximately 6 weeks after the general testing, we invited back the high scorers for some additional testing, both cognitive and noncognitive. These students were those who had ranked the highest on math and/or science. Accordingly, 35 boys were called back for further testing, and all of them came. Ten girls who ranked highest among the girls were also asked to come, and eight of them came. Since the girls' scores on the competitive measures (SAT-M, M-I, and Science) were not as high as the boys', nor even as high as some boys who were not invited back for reasons of space and material, the subsequent analyses of the "high group" will refer to the 35 boys who were tested a second time.

CONCOMITANTS OF MATHEMATICAL PRECOCITY

There is some corollary information available about the total group. With the registration materials was sent a questionnaire, which the students were requested to bring with them on the day of the testing. Of the total of 450 contestants, 416, or 92%, returned the questionnaire. On none of the three test measures (SAT-M, M-I, or Science) was there a difference even approaching significance between those who returned questionnaires and those who did not; thus information gleaned from the returned questionnaires may be assumed to be representative of the whole group. Also on the day of the testing a rearranged checklist of occupations from Holland's (1965) Vocational *Preference Inventory* (VPI) was administered to all the students. In terms of the total group (i.e., not breaking down by sex or grade) and the high group, it will be fruitful to look at the VPI and at three items from the questionnaire: the students' reported liking for school; the level of education attained by the father and the mother; and the sibling position of the student.

Liking for School

On the questionnaire, the students were asked, "How would you describe your liking for school?" There were four options listed: strong liking, fair liking, slight liking, and dislike. The responses were purposely weighted in *favor* of a positive response (three "liking" categories and only one "dislike" category) so that a negative response would have to be intentional. As can be seen in Figure 1, the mean scores for the total groups *increase* monotonically as the degree of reported liking for school decreases for all three tests. The number of students responding in the first two categories is much larger than in the second two, which complicates the interpretation, and the picture in science is further confounded by the overlapping math and science people. But generally the trend is quite clear: gifted seventh- and eighth-grade students who were advanced enough to get the high scores on these college-level tests reported less "liking for school" than gifted students who do not do as well on such difficult tests.

If we look at the high group in comparison with all boys, the picture is even clearer. For all boys, the percentage reporting each of the four categories, strong liking to disliking, were: 27.9, 55.1, 15.4, and 1.6; for the high group, the comparable figures were respectively: 17.7, 55.19, 20.6, and 5.9. Also striking is the fact that of the total of four boys who reported dislike for school, two were in the high group, with SAT-M scores of 780 and 740 and M-I scores of 720 and 630.

This result confirmed expectations derived from earlier informal discussions with some of the precocious students who had been identified prior to the general testing. They reported a disillusionment with school in particular and academic pursuits in general. The boredom and frustration generated by their school situation were expressed in different ways by different students, as might be expected, but it was frequently present.

A completely different interpretation of these data is possible if one were so disposed, i.e., that bright, precocious students are on the average more maladjusted than "normal" ones. In a

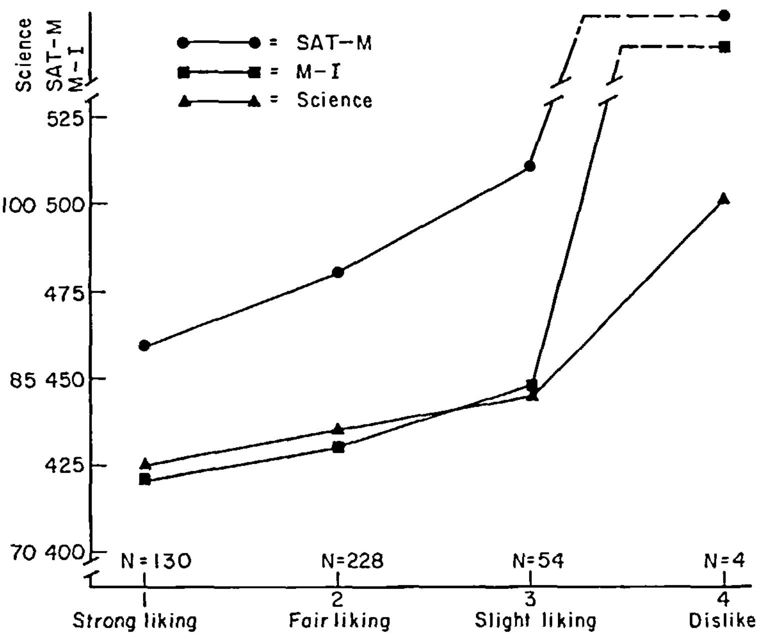

Figure 1. Mean scores of all students on three measures for degrees of reported liking for school.

sense, this would be true: *Situationally*, some of them do seem to be maladjusted. But this may well be a healthy reaction to intolerable circumstances. *Personally*, they appear to be a well-adjusted group. The California Psychological Inventory (CPI—Gough, 1964) was administered to the high group when they were called back for retesting. Although there is considerable variance in the sample, there is no indication that mathematically precocious boys are especially subject to disabling inter- or intrapersonal weaknesses. Thus the "maladjustment" interpretation is implausible if its meaning is anything other than situational maladjustment.

In the light of the tentatively confirmed expectation of lack of enthusiasm for school, it is indeed difficult to value highly some of the frequently heard arguments against significant restructuring of these students' educations because of the potential harm to their social and emotional development. A number of studies suggest strongly that such a fear is unfounded (e.g., Oden, 1968; Pressey, 1949). But the obverse of that concern is one voiced much less often: What is the potential harm to the social and emotional development of these students if they are required to remain in an unstimulating, hence frustrating, environment? It may be great, as we have suggested elsewhere in case-study fashion (Keating & Stanley, 1972) and as these data tentatively indicate.

Some of the best students, however, did report "strong liking" for school. Being placed in an educational situation far below one's capacity does not inevitably result in less liking for school. This is merely another expression of the fact that even though this high group is homogeneous on a number of variables, it is quite heterogeneous on others such as temperament (i.e., those not used for selection). In the second large screening session to be held soon, the questions will be rephrased slightly in an attempt to separate liking for school and liking for mathematics.

Sibling Position

It can be anticipated that birth order will be associated with most data about individuals one might conceivably wish to collect, including aptitude and achievement (Breland, 1972; Lunneborg, 1971). But, just as there was an unexpected sex-difference in these data, there was an unanticipated *lack* of any differences due to sibling position in the total group. The three virtually straight lines in Figure 2 illustrate this point well.

Standing in contrast to this flat profile is the distinct pattern within the high group, as shown in Figure 3. On SAT-M and M-I, those in the high group born second score higher than only children, firstborns, or those born third or later (but none of these differences are statistically significant). This observation relates to a finding by McGurk and Lewis (1972), who report an effect on dependency behavior for those born second. As they suggest, it is a "phenomenon in search of an explanation." It is also a phenomenon showing up inconsistently and weakly enough in these data to militate against excessive speculation.

Parents' Education

If in our study the unexpected and the unusual has been the rule, then the association of the parents' education and children's test scores is the exception: It followed the often replicated pattern to perfection. In the total group, the higher the reported educational level of the parents, the higher were the mean test scores of their children on all three measures (SAT-M, M-I, and Science). Indeed, the relationship was monotonic for both mother's and father's education on SAT-M and nearly so on M-I and Science. Figure 4 depicts this relationship for father's education, and Figure 5 for mother's education.

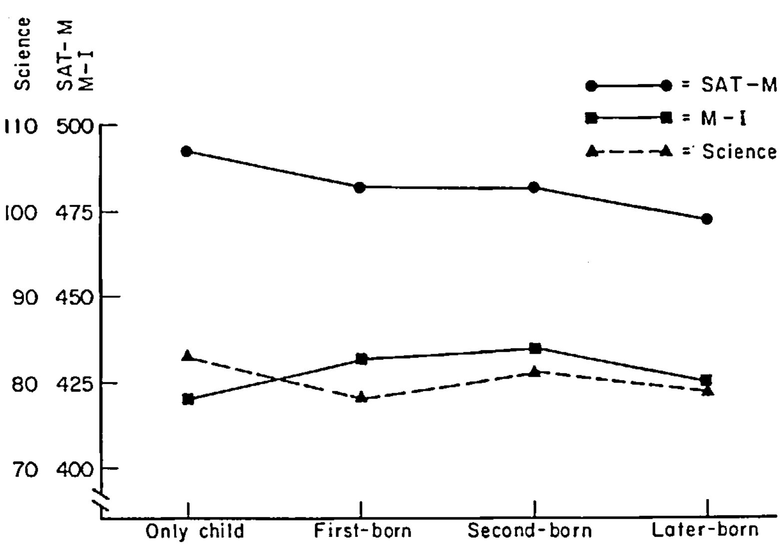

Figure 2. Mean scores of all students on three measures by sibling position.

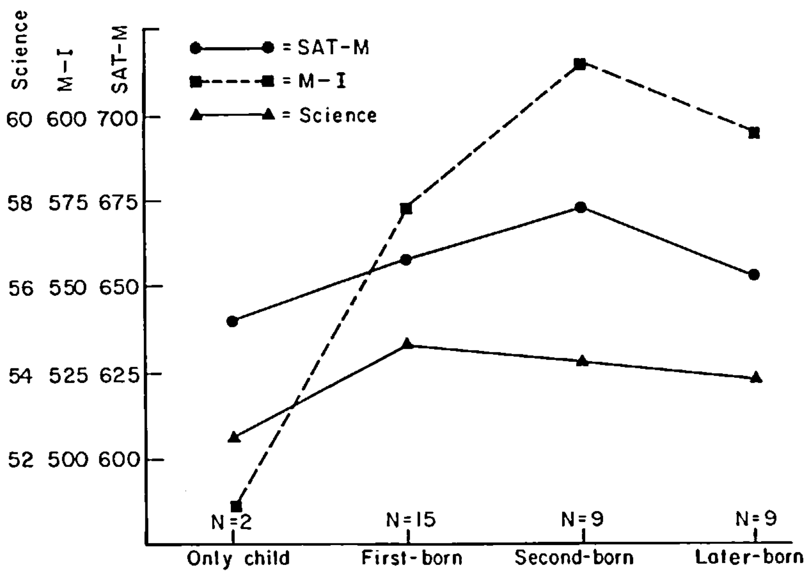

Figure 3. Mean scores of the high group on three measures by sibling position.

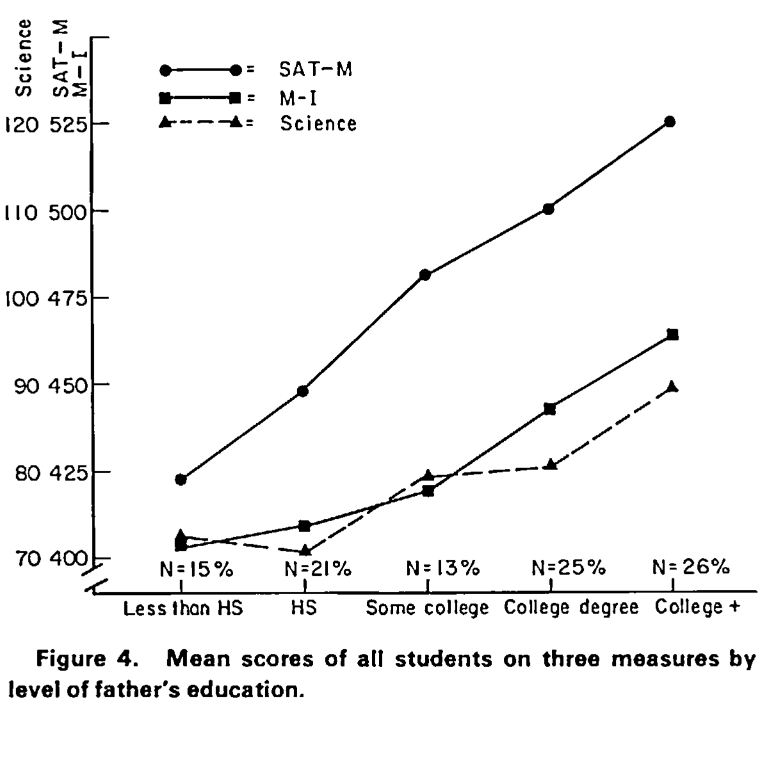

Figure 4. Mean scores of all students on three measures by level of father's education.

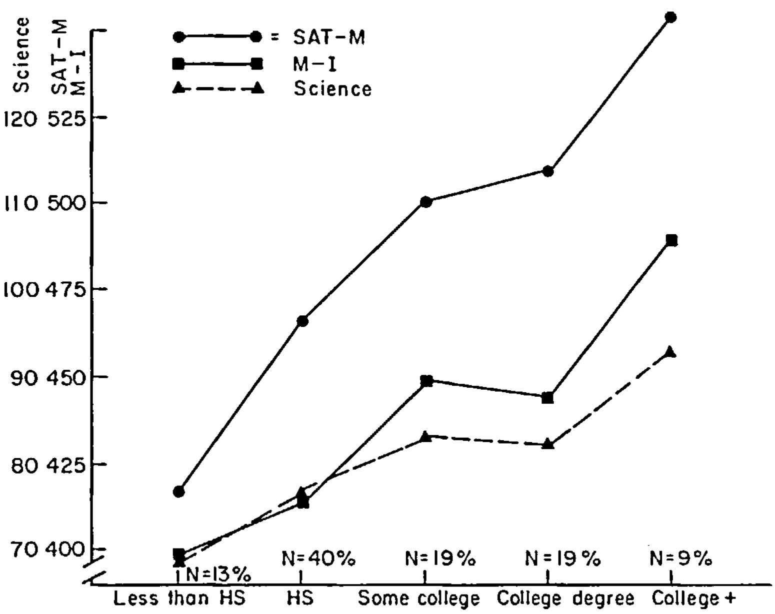

Figure 5. Mean scores of all students on three measures by level of mother's education.

This result is neither surprising nor, unfortunately, especially informative. Does this merely reflect the finding that bright parents tend to have bright children? Or does it say something about the relatively different types of experience to which the children with the higher test scores might have been exposed? Or both? These data offer no solution to that dilemma.

Slightly more informative are the figures for the high group. The pattern is similar within that group, with 47% of the fathers and 28% of the mothers having education beyond a college degree. *But*, and these are noteworthy exceptions, 29% of the fathers and 51% of the mothers had less than a college degree. Of the fathers, in fact, 12% were reported not to have even a high school diploma. Thus, while both the fathers and the mothers of the high group were reported to be significantly better educated than the parents of the remainder of the contestants, there was a substantial amount of this high level talent coming from homes where the parents were *not* professionals.

The contestants reported the highest educational levels attained by the father and the mother. These were coded in the following fashion: 1—less than high school diploma; 2—high school diploma; 3—some college; 4—college degree; and 5—courses beyond the college degree. Interestingly, the larger difference between the high group and the remainder group was in terms of the mother's education, with a mean of 3.56 for mothers of the high group and 2.57 for mothers of the remainder group ($p < 0.001$). For fathers the means were 3.91 and 3.24, respectively ($0.001 < p < 0.01$). This is partially attributable, of course, to the relatively high correlation ($r = 0.62$, a correlation significantly different from 0 at $p < 0.001$) between the reported educational level of the fathers and mothers within the high group. In Figure

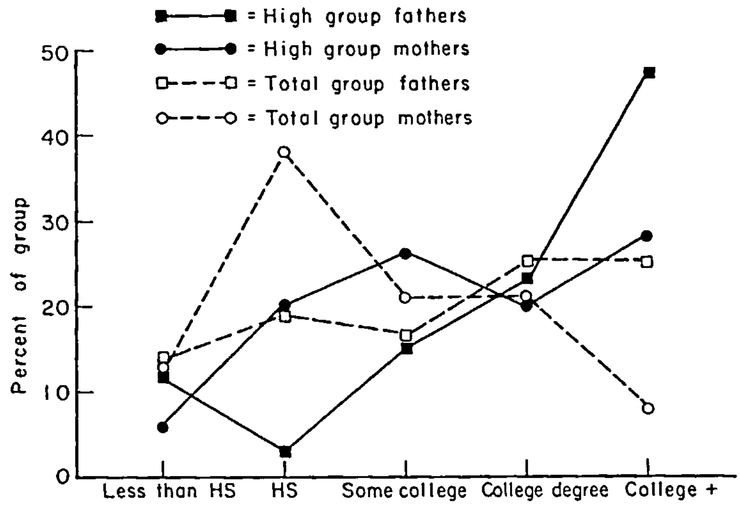

Figure 6. Percent of high group and total group parents at different educational levels.

6 the percentages of the comparison group for the highs and the total group for each educational level are shown for both mothers and fathers. There is a fairly consistent pattern in the differences between the groups for mothers and fathers.

Vocational Interest

All the contestants were requested to complete the checklist taken from the VPI. The resulting codes are thus somewhat more unreliable than they would presumably be if the longer (but for our purposes somewhat less appropriate)

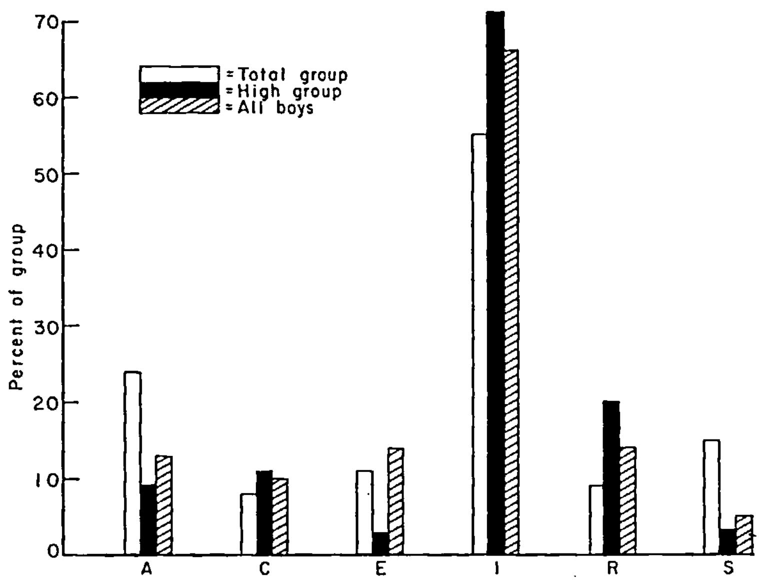

Figure 7. Percent of total group, all boys, and high group with each VPI code (1st letter).

instrument were used. For these and other reasons, only the first letter of the Holland three-letter code will be discussed in this paper.

In Figure 7, the percentages of three groups (total, highs, and boys) having each of the six code letters as the first letter are shown. The codes are (A)rtistic, (C)onventional, (E)nterprising, (I)nvestigative, (R)ealistic, and (S)ocial. The percentages across codes, incidentally, sum to over 100%, since it is possible to have more than one first letter code (i.e., a tie for first letter). Eighteen percent of all boys and 11% of the high group had more than one first letter.

Immediately striking is the height of the Investigative columns. Over 66% of the boys overall, and 71% of the boys in the high group had "I" as at least one of their first letters. This is not surprising, since all of the "I" occupations are science oriented, whereas only a few under the other five codes are even remotely related to science. There is a further confounding in that the status and educational level of the "I" occupations (exclusively professional) is unmatched by any other code.

It is, nevertheless, a result which confirms some earlier speculations about the interests of students like these. They appear above all to be interested in finding out things, discovering things, learning things. The absence of any difference in the relative percentages of the high-group boys and all boys on the first letter of the Holland code would seem to indicate that the interests are the same, but the high boys have for some reason acted upon those interests more effectively.

Interestingly enough, as Figure 8 portrays, the mean scores for boys having the different letters as the first one of their Holland code follow a consistent pattern. Despite the large number who had "I" as their first letter, which would tend to depress the mean score for "I" toward the grand mean of the boys, its mean is highest in science and second highest in both SAT-M and M-I. Those boys having (A)rtistic and

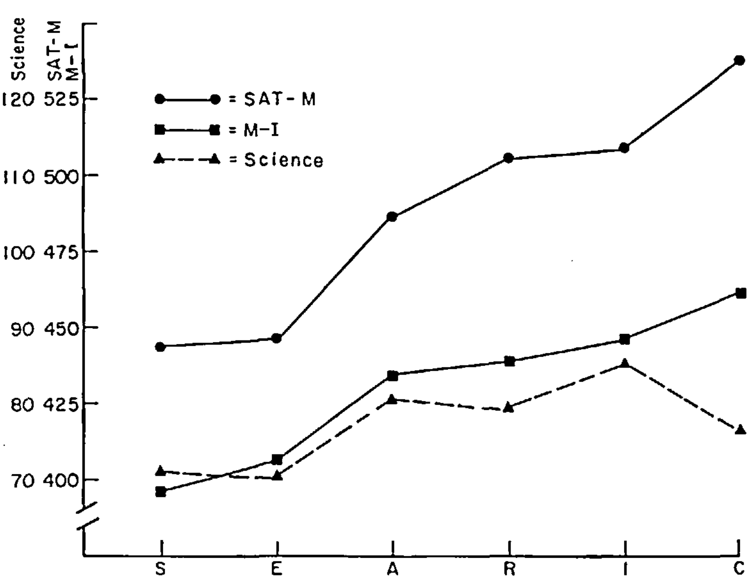

Figure 8. Scores of all boys on three measures by first letter of VPI code.

(E)nterprising as their first letter were consistently the lowest on all tests. This indicates that precocity is more than just a high level of ability, that it is a subtle blend of ability, interest, personality, and probably other factors as well. These other factors may well include evaluative attitudes, as we shall see later.

THE PSYCHOMETRY OF MATHEMATICAL PRECOCITY

If the best of the students on the basis of the general competition were to be facilitated educationally, and if more was to be learned about the pattern of their abilities, it was essential that they be tested more extensively and intensively than had already been done. Accordingly, as mentioned above, 35 boys (the "high group") were invited back for further testing. These included: Scholastic Aptitude Test-Verbal (SAT-V), which was the most crucial and which all in the high group took; SAT-M for the few science-only qualifiers; Science (Form A only) for those math-only qualifiers; Mathematics Achievement, Level II (M-II) for nine boys who had done especially well on M-I; Raven's Progressive Matrices, a test of nonverbal reasoning ability; Terman's Concept Mastery Test of vocabulary and verbal reasoning; Bennett's Mechanical Comprehension Test; and the Revised Minnesota Paper Form Board Test. Obviously, not all the high group could take all the tests even in a day-long testing session. A second retesting has been conducted, and all the available scores and summary statistics on the tests enumerated above are listed in Table 3.

There are a number of observations to be made concerning the means in Table 3. The SAT-M mean of 660 was "expected" since it was the major criterion for an invitation to come to the retesting. It is actually lowered by the fact that several contestants who qualified on the basis of their science score, but who had relatively low SAT-M scores, are included.

The SAT-V mean score of 546 for this group is about 0.4 of a standard deviation lower than the mean SAT-M score, based on the national sample of male high school seniors. This again is to be understood partially in terms of the selective criteria and the expected regression toward the mean.[2] But it may also be indicative of something a bit deeper. Verbal precocity may be rather rarer than the quantitative variety. Mathematics may be, in some psychologically meaningful sense, a closed system, whereas vocabulary and perhaps verbal reasoning may be somehow "openended," more dependent on accumulated experience. This will be subject to investigation in the future, especially in the analysis of results from the second large screening competition, which included a mathematical and a verbal section.

It has been suggested to us by a number of teachers that the use of such rigorous tests for seventh and eighth graders is perhaps useless or harmful or both. What is the rationale for using these high-level tests at such an early age?

There are really three related answers to that question, one pragmatic and the other two somewhat more theoretical. As we have reported elsewhere (Keating & Stanley, 1972), the two "radical accelerates" at Johns Hopkins were evaluated on the basis of the tests normally given to prospective freshmen. Their successful performance on them was in fact indicative of their

[2] Whenever a group is selected on the basis of a test score criterion above or below the population mean, the mean score of the select group on a subsequent measure will be closer to the population mean than on the original measure. This statistical phenomenon is known as "regression toward the [population] mean."

TABLE 3
SCORES OF HIGH SCORERS[a] FROM SCIENCE FAIR COMPETITION ON VARIOUS COGNITIVE MEASURES

Student boys	Scholastic Aptitude Test		CEEB Math Achievement		Science[b]		RPM[c]		CMT[d]		MCT[e]		RMPFB[f]	
	Math	Verbal	Level I	Level II	A	B	A	II	I	II	AA	CC	MA	MB
1	790	560	770	750	54	52	59	33	26	49	–	34	57	–
2	780	620	720	720	60	66	59	31	43	48	53	50	48	52
3	740	620	660	670	51	–	59	33	29	51	–	–	–	–
4	740	460	630	620	38	–	57	33	14	25	20	27	61	59
5	730	560	620	610	46	54	55	30	31	33	38	32	–	–
6	710	530	730	800	69	68	57	32	19	31	54	45	49	51
7	710	640	640	690	53	56	58	29	33	52	35	22	59	58
8	690	550	590	–	56	–	–	–	17	48	48	40	–	–
9	690	450	520	–	43	36	58	35	26	39	32	31	40	–
10	680	670	720	690	43	–	55	26	34	31	–	18	42	–
11	680	620	570	–	63	65	56	35	46	51	48	37	–	–
12	680	540	610	–	48	–	58	33	23	38	41	39	–	–
13	680	450	500	–	42	–	51	25	21	29	39	33	–	–
14	680	500	510	–	44	–	–	–	28	20	41	23	–	–
15	670	650	660	620	64	61	55	33	19	37	50	11	48	50
16	670	570	610	710	60	66	59	29	34	33	50	46	–	–
17	670	590	600	–	59	59	54	33	26	50	53	52	–	–
18	660	490	520	–	47	–	–	–	8	27	42	34	–	–
19	660	460	520	–	54	–	60	30	8	27	50	32	37	43
20	660	460	580	–	40	10	58	32	−1	22	39	30	–	–
21	660	420	530	–	44	49	60	29	12	29	40	34	–	–
22	660	310	600	–	39	–	56	31	9	29	41	21	–	–
23	650	540	560	–	46	–	56	30	–	–	43	36	–	–
24	640	580	610	–	57	55	58	29	6	35	–	–	–	–
25	640	400	510	–	39	–	–	–	12	28	24	34	–	–
26	630	580	500	–	53	63	59	28	46	26	53	47	52	–
27	620	740	520	–	59	59	58	31	48	47	20	27	52	58
28	620	530	640	–	39	–	53	27	2	28	12	18	–	–
29	610	530	520	–	57	55	59	34	6	35	56	52	–	–
30	600	600	–	–	62	64	56	30	37	42	41	36	51	41
31	600	550	520	–	58	61	–	–	22	37	38	36	48	–
32	590	550	–	–	56	54	57	31	14	33	45	36	–	–
33	560	620	540	–	57	57	–	–	45	40	47	40	–	–
34	530	550	470	–	66	64	59	33	19	30	49	46	–	–
35	520	630	510	–	56	57	51	28	38	48	–	–	–	–
N	35	35	33	10	35	22	29	29	34	34	30	32	13	8
Mean	660	546	585	688	52	58	57	31	24	36	41	34	50	52
S.d.	60	86	77	61	9	8	2	3	14	9	11	10	7	7

[a]The 35 highest scoring competitors on math and/or science, all boys, ranked in order of SAT-M score. Students 29 thru 35 qualified on basis of *science* score.

[b]Sequential Tests of Educational Progress, Series II (STEP II), Forms 1A and 1B. (Educational Testing Service).

[c]Raven's progressive Matrices, Sets A, B, C, D, E (listed as "A" in table) and Set II ("II" in table).

[d]Concept Mastery Test, Parts I and II (Psychological Corporation).

[e]Bennett's Mechanical Comprehension Test, Forms AA and CC (Psychological Corporation).

[f]Revised Minnesota Paper Form, Board, Forms MA and MB (Psychological Corporation).

ability to succeed admirably at college-level academic tasks. In simple language, the tests work. The continuing predictive success of these tests for this purpose has justified their continued use.

The second point has to do with the unbounded nature of the distribution beyond the 99th percentile. All *z*-scores of 2.33 or greater are included in the traditionally reported 99th percentile, beyond which age-in-grade testing simply does not make distinctions or, more characteristically, does not possess enough ceiling to make such distinctions. The top 1% of seventh and eighth graders comprises a group from which those capable of such advanced and rapid learning must be identified. Although further in-grade testing is therefore unproductive, evaluation with college-level tests yields important information.

Third, these high-level tests are more appropriate than in-grade tests because the former almost necessarily tap higher-level abilities than the latter. The following example demonstrates this distinction. A student who has learned the formula for finding the area of a right triangle will apply it readily when presented with that type of problem. The ability required to solve the problem in that circumstance is of a lower order than the ability required when the same task is presented to a student who has not yet learned the formula directly and must deduce it. Similar distinctions may be drawn between the high school senior taking the SAT and the seventh or eighth grader doing so. The "stimulus" is objectively the same, but it is a quite different task in terms of the individual. This is discussed more fully elsewhere by the author (Keating, 1975, in press).

The small size of the target group is relative, not absolute. A conservative estimate of the number of junior high school students (seventh, eighth, and ninth graders) nationwide fitting into this exclusive category is 2,500 to 3,000. The principles and practices being developed in this study would presumably extend beyond even this limited group, however.

The only other test taken by all 35 boys was science, again including those who had chosen not to take it the first time around. The mean score on Form A for this high group is 52 (of a possible 75), which is the 80th percentile of college sophomores tested in the spring. Several of these students, however, demonstrated a grasp of general science knowledge better than 95% of the comparison group of college students.

Most of the 35 students have taken the Raven's Progressive Matrices, Sets A-B-C-D-E and Set II. Their mean score on Set II is 31 (of a possible 36) and the median is 30. These are 2.5 and 2.3 standard deviations above the mean of British university students. The British standardization may be slightly suspect (it was not an excellent sample of subjects or situations), but clearly these students excel at nonverbal reasoning ability by any standard. Two of the students (one of them a seventh grader) missed a perfect score by only one item on this extremely difficult test, and several more by but two or three items.

Terman's Concept Mastery Test, which he devised to measure the adult intellectual stature of his gifted group (Terman, 1947), consists of two parts. The first is a pure vocabulary test, and the second is a test of verbal reasoning by use of incomplete analogies. Further insight into the strengths and weaknesses of these mathematically precocious youngsters is gained from comparing scores on the two parts. With only two exceptions, these students had corrected raw scores on the second part which were higher than those on the

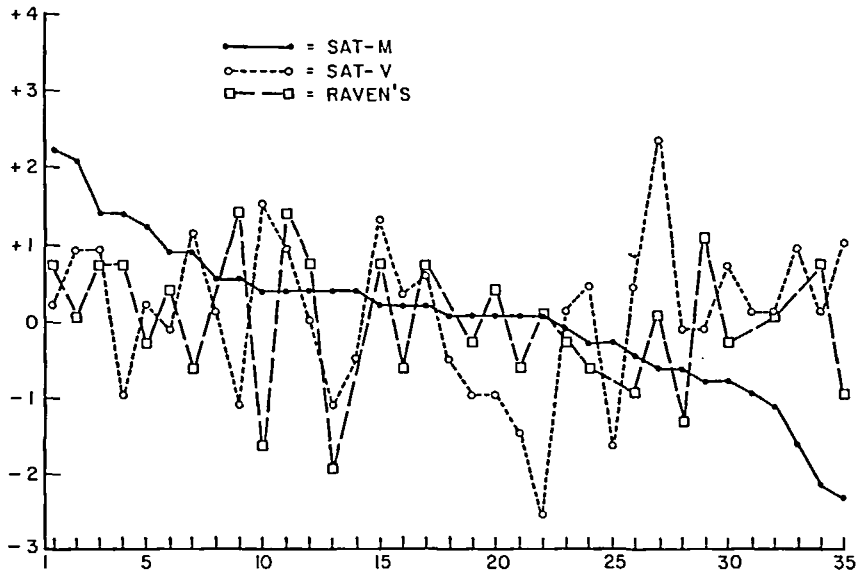

Figure 9. Z-scores on three measures for each student in high group.

first part, even though there are 115 items on part one and only 75 on part two. Thus as a group, and as individuals, they are better at reasoning at an early age, even verbal reasoning, than at tasks requiring purely verbal aptitude.

Little in the way of general conclusions can be drawn from the tests of mechanical comprehension and spatial ability. On the average, these students score at the 36th percentile of engineering freshmen on mechanical comprehension and at the 55th percentile of a similar group on spatial relations. There is great variability in the scores, however. Several students who are good at spatial relations are almost totally lacking in mechanical comprehension. The converse may also be true, but not enough of the students took the spatial relations test for us to know.

This does, however, raise an important point which requires further elaboration: the inter- and intra-individual variability of the group mentioned earlier. Figure 9 depicts the latter quite vividly. The students were ranked on SAT-M and their z-scores within this group computed and plotted. It is the solid line going from top to bottom, left to right. The same procedure was followed for their SAT-V and RPM scores. The maze of criss-crossing dotted lines is the result. What the figure shows is that there is a decided lack of predictability of rank within the group from any of these three tests to any of the others. Some of these students are far better than the others on some tasks and poorer on others.

In a more conventional format, Figure 10 shows the lack of relationship within this group between SAT-M and SAT-V. The correlation equals -0.08. The correlation between SAT-M and RPM is 0.20. Neither is significantly different from zero. This admittedly borders on the restriction of range fallacy, i.e., within such a select group a high correlation between the criterion and other aptitude measures is not to be anticipated. But one additional analysis lends support to the intuitive perception of great variability. The mean absolute z-score differences (i.e., disregarding directionality) between SAT-M and SAT-V and between SAT-M and RPM for this group are 1.18 and 0.96, respectively. Thus the average difference in both cases is approximately a standard deviation. This also reflects the variability of the high group.

In addition to these tests, the students were also given the Allport-Vernon-Lindzey Study of Values (AVL) and the California Psychological Inventory (CPI). The CPI results are examined in detail by Weiss, Haier, and Keating (in Chapter 7 of *MT:D*[3]). The AVL was used to determine whether the evaluative attitudes of these students were similar to those of creative scientists, on the assumption that this might be a good indication of their potential for later achievement. MacKinnon (1962) reported that the classic pattern of eminent *adult* scientists is Theoretical-Aesthetic. These values appear to be conflicting, but the combination seems conducive to creativity in science. A desire for knowledge and understanding for their own sake is the *sine qua non*, but this must be mediated by an appreciation of form and harmony (e.g., the "elegant" proof in mathematics).

The theoretical value is prevalent in this sample. Twenty of the 35 students had it as their first (highest) value (on the basis of male college population percentiles, not raw scores), and 7 of the remaining 15 had it as the second. This result, taken together with the information from the Holland checklist, confirms the common sense notion that these precocious youngsters must possess a strong intrinsic interest in learning and knowing. Thus one of the first (and perhaps major) noncognitive, personological ingredients for eventual creative achievement in the sciences seems to be present in abundance.

There is also some evidence to validate the importance of the theoretical scale. The science scores seemed to be a good indicator of independent interest in science. The raw score on the theoretical scale was correlated at $r=0.55$ with the science score within the high group. It correlated more highly with the science score than any cognitive measure except SAT-V.

The aesthetic value, however, did not fare too well in this group. Only nine people had it as either their first or second highest value. Five who had theoretical as their first value had aesthetic as their second, showing the "classic" creative scientist pattern. This may indicate several possibilities. First, it may be that the theoretical-aesthetic pattern is rare in the population and that our finding represents a marked positive deviation from the base rate. Research on this specific point has not, to our knowledge, been done. Thus, the expectation should perhaps be lower, and finding even these few is significant.

Second, it may be that the adult creative scientists who were studied developed their evaluative attitudes when they were much older than these students. Perhaps very few 12- to 14-year olds show much of an aesthetic evaluative attitude, and the theoretical-aesthetic creative scientists did not when they were that age. This suggests two further possibilities. One is that this is a maturational-developmental sequence which may emerge in any event. The other is that certain adolescent or post-adolescent experiences of the creative scientists engendered this value construct, and experiences similar to these need to be made available to these students. There is some evidence that this is true for a more diverse group. Huntley (1965) reported a significant increase on the aesthetic value for science majors over 4 years of college. One may

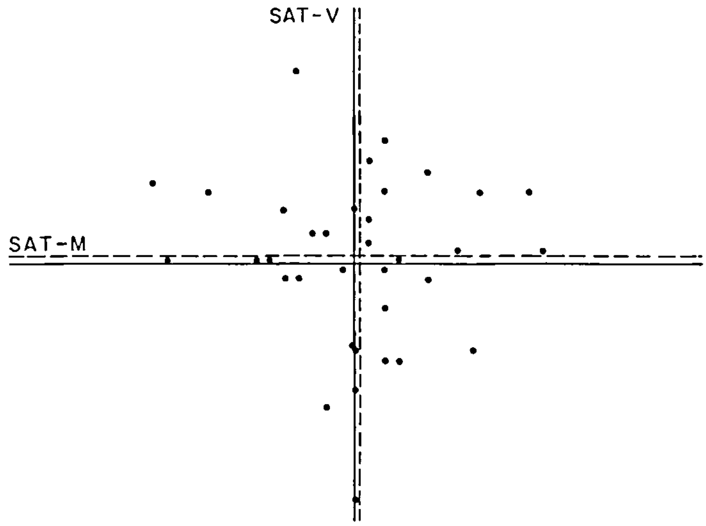

Figure 10. Scatterplot of z-scores on SAT-M and SAT-V in high group (r = -0.08).

be concerned to learn, however, that fathers of boys in the high group do not score much higher on the aesthetic scale than their sons do.

A third point to be considered is much more speculative but also more disturbing. McCurdy (1957) reviewed the childhood patterns of 20 geniuses. One strong pattern that emerged is a long and continuous contact of the genius-to-be with adults on a nearly equal basis from a very early age. Peer-group interaction of the sort prized above all else in today's schools is conspicuously absent. It is not a simple task to construe the relationship between this and creativity, or even an aesthetic attitude, but perhaps one exists. This would constitute a historical-cultural interaction of childrearing practices and the frequency of creative individuals. It would mean that by constructing our schools as we have and by making such education universal and compulsory from age 5 or 6 onward, we may have in fact unknowingly militated against the development of creativity. To speculate further involves diffcult and complex questions beyond the scope of this paper.

EDUCABILITY OF MATHEMATICAL PRECOCITY

The screening of gifted seventh and eighth graders by means of a general test competition turned up a number of youngsters who are mathematically precocious. The analyses of the total group of contestants and the high group sought to answer important questions about the nature and development of their talent. The results suggested the possibility that "precocity" might be educable (in the root sense of "leading or drawing our from") with a highly select group of students and effective techniques.

On hand from a related study were the names and scores on the Academic Promise Tests (APT) of 392 sixth graders who had been nominated by their teachers or principals as gifted (10 students from each of 40 schools). The scores were on four subscales: number

(N), verbal (V), abstract reasoning (AR), and language usage (LU). On the basis of the following criteria, approximately 30 students were invited to participate in an algebra class to be conducted on Saturdays during the summer but which would rely primarily on independent study during the week. The student had to have a 99th percentile score for the sixth grade national norms on "N" and also a 99th percentile on either "V" or "AR."

The class began June 1972 with 19 students, 16 of whom came from the pool of 392 students. Three special subjects were also offered the opportunity and took it, including a remarkably able boy who had just completed the third grade.

CONCLUSION

As is inevitably the case when one confronts an uncharted area of investigation, we are more aware of what we do not know than what we do know. Although that is true of our study of mathematical precocity, we have learned a great deal.

Chapanis (1971) discusses several "levels of explanation" for behavioral data. We have begun to get a good grasp on the first level, which concerns practical applications. We have developed and will be refining adequate methods for identification of mathematically precocious youth, and our armamentarium of facilitative techniques is large and growing, as the following paper will demonstrate.

At the deeper level of basic understanding, however, we are still groping. We have some leads in the correlational data we have collected and some speculations about them. Fortunately, our ability to use new methods and approaches is frequently not limited by our understanding of the rules which underlie them.

References

Allport, G., Vernon, P. E.; Lindzey, G. *Manual for the study of values* (3rd ed.). Boston: Houghton Mifflin, 1970.

Breland, H. M. Birth order, family configuration, and verbal achievement. *Research Bulletin*, 1972, Educational Testing Service, RB-72-47.

Chapanis, A. The search for relevance in applied research. In W. T. Singleton, J. G. Fox, and D. Whitfield (Eds.). *Measurement of man at work*. London: Taylor and Francis Ltd., 1971.

Cox, C. M. *The early mental traits of three hundred geniuses*. Vol. 2., *Genetic studies of genius*. L. M. Terman (Ed.). Stanford, Calif.: Stanford University Press, 1926.

Galton, F. *Natural inheritance*. New York: MacMillan, 1889.

Gottesman, I. I. Genetic aspects of intelligent behavior. In Ellis (Ed.). *Handbook of mental deficiency: Psychological theory and research*. New York: McGraw-Hill, 1963.

Gough, H. G. *Manual for the California Psychological Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1964.

Holland, J. L. *The Vocational Preference Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1965.

Holland, J. L. *The self directed search*. Palo Alto, Calif.: Consulting Psychologists Press, 1970.

Huntley, C. W. Changes in Study of Values scores during the four years of college. *Genetic Psychology Monographs* 1965, *71*, 349-383.

Keating, D. P. Testing those in the top percentiles. *Exceptional Children, 1975, in press.*

Keating, D. P., & Stanley, J. C. Extreme measures for the exceptionally gifted in mathematics and science. *Educational Researcher*, 1972, *1* (9), 3-7.

Lehman, H. C. *Age and achievement*. Princeton, N.J.: Princeton University Press, 1953.

Lunneborg, P. W. Birth order and sex of sibling effects on intellectual abilities. *Journal of Consulting and Clinical Psychology*, 1971, *37* (3), 445.

MacKinnon, D. W. The nature and nurture of creative talent. *American Psychologist*, 1962, *17* (7), 484-495.

McCurdy, H. S. Childhood pattern of genius. *Journal of the Elisha Mitchell Scientific Society*, 1957, *73*, 448-462.

McGurk, H., & Lewis, M. Birth order: A phenomenon in search of an explanation. *Research Bulletin*, 1972, Educational Testing Service. RB-72-20.

Pressey, S. L. *Educational acceleration: Appraisal and basic problems*. Bureau of Educational Research Monographs, No. 31. Columbus: Ohio State University Press, 1949.

Terman, L. M., & Oden, M. H. *The gifted child grows up*, Vol. 4. *Genetic studies of genius*, L. M. Terman (Ed.). Stanford, Calif.: Stanford University Press, 1947.