DOCUMENT RESUME

ED 080 583 TM 003 100

AUTHOR	Skager, Rodney; Fitz-Gibbon, Carol
TITLE	Mentally Gifted Disadvantaged Students: An Investigation of Methods of Identification, Including the Use of "Culture Fair" Tests, at the Eighth Grade Level. Final Report.
INSTITUTION	California Univ., Los Angeles.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C. Regional Research Program.
BUREAU NO	BR-1-I-057
PUB DATE	Apr 72
CONTRACT	OEC-9-71-0039(057)
NOTE	79p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Ability Identification; *Cognitive Tests; Cultural Factors; Culture Free Tests; Disadvantaged Youth; Gifted; Grade 8; Group Tests; *Performance Criteria; Questionnaires; *Screening Tests; Student Attitudes; *Student Testing; *Superior Students; Test Results
IDENTIFIERS	Advanced Progressive Matrices; California Achievement Test; California Test of Mental Maturity; Standard Progressive Matrices; WISC

ABSTRACT

The aim of this investigation was to develop a justifiable and efficient procedure for locating the top two percent of the eighth grade students in an inner-city junior high school. The WISC performance IQ was chosen as the criterion measure. Four screening measures were used to select students for the WISC: an achievement test (the California Achievement Test 1970 reading and mathematics), a conventional IQ test (the California Test of Mental Maturity short form 1963), a "culture fair" test (Raven's Standard Progressive Matrices) and teacher nominations. Students scoring in the top two percent of the sample on one or more of these measures took the Advanced Progressive Matrices and the WISC. The top nine students on the WISC performance scale were designated mentally gifted. Since four of the nine gifted students were not nominated by any teacher, the use of teacher nominations is not recommended except as a supplement to objective tests. A procedure involving the use of the SPM followed by the APM is recommended for selecting students for the WISC. This procedure would have identified eight of the nine gifted students in this study, and would have required the administration of only 14 WISCs. (Author)

Final Report

Project No. 1-I-057
Contract No. OEC-9-71-0039(057)

Rodney W. Skager, Carol Fitz-Gibbon
University of California at Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90024

MENTALLY GIFTED DISADVANTAGED STUDENTS:

an investigation of methods of identification, including the use of "culture fair" tests, at the eighth grade level.

April 1972

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education
Regional Research

Final Report

Project No. 1-I-057
Contract No. OEC-9-71-0039(057)

Rodney W. Skager, Carol Fitz-Gibbon
University of California at Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90024

MENTALLY GIFTED DISADVANTAGED STUDENTS:

an investigation of methods of identification, including the use of "culture fair" tests, at the eighth grade level.

April 1972

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education
Regional Research

AUTHOR'S ABSTRACT

The aim of this investigation was to develop a justifiable and efficient procedure for locating the top two percent of the eighth grade students in an inner-city junior high school. The WISC performance IQ was chosen as the criterion measure.

Four screening measures were used to select students for the WISC: an achievement test (the California Achievement Test 1970 reading and mathematics), a conventional IQ test (the California Test of Mental Maturity short form 1963), a "culture fair" test (Raven's Standard Progressive Matrices) and teacher nominations. Students scoring in the top two percent of the sample on one or more of these measures took the Advanced Progressive Matrices and the WISC. The top nine students on the WISC performance scale were designated mentally gifted.

Since four of the nine gifted students were not nominated by any teacher, the use of teacher nominations is not recommended except as a supplement to objective tests.

A procedure involving the use of the SPM followed by the APM is recommended for selecting students for the WISC. This procedure would have identified eight of the nine gifted students in this study and would have required the administration of only 14 WISCs.

FINAL REPORT

Project No. 1-I-057
Contract No. OEC-9-71-0039(057)

MENTALLY GIFTED DISADVANTAGED STUDENTS:

an investigation of methods of identification,
including the use of "culture fair" tests, at
the eighth grade level.

Rodney W. Skager
Project Director

Carol Fitz-Gibbon
Author

University of California at Los Angeles
Los Angeles, California

April 1972

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

**U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE**

Office of Education
Regional Research Program

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF FIGURES

CHAPTER I

INTRODUCTION

It is hoped that this investigation of methods of identifying mentally gifted students in an inner-city junior high school will assist those who are responsible for this procedure in the schools.

Recent legislation in California, permitting the use of special criteria in the identification of mentally gifted disadvantaged students, has greatly facilitated the establishment of gifted programs in inner-city schools. This enables more minority students to benefit from the use of gifted funds which previously were used almost exclusively in advantaged schools. With the increasing demand for minority personnel at high professional levels, it is essential that talent be located early and enabled to flourish.

The special criteria in California

Prior to the 1969 passage of an amendment to the California State Code, the criteria which a secondary school student had to meet to be placed in a gifted program were scores above the 98th percentile point on both a group intelligence test and on a standardized achievement test. These are the criteria specified by section 3821 of title 5 of the California State Code. Since very few students in disadvantaged schools were likely to meet these criteria, gifted programs rarely existed in such schools.

The 1969 amendment allows students who are culturally disadvantaged and underachieving to be identified as "mentally gifted" under separate criteria specified by section 3822 of title 5 of the California State Code. According to section 3822 a student is considered mentally gifted if "he may be expected, within a reasonable time and with appropriate curricular modifications, to perform in school at a level equivalent to that of the mentally gifted minors identified pursuant to section 3821."

The legislation continued by stating that this judgment shall be based on characteristics such as "precocious development", "unusual resourcefulness" or "outstanding achievements". The only objective measures which are specified as possible criteria are "scores at or above the 98th percentile on nonverbal (performance) scores of an individual intelligence test".

Finally, there is a clause which states that the number of students identified under the special criteria shall not exceed two percent of the culturally disadvantaged pupils within a school district.

Assumptions underlying the design of the study

One of the first requirements of a selection procedure is that each student is afforded an equal opportunity to be considered. This requires some kind of initial screening of all students in a grade level, about 400 students in an average sized junior high school. It is doubtful that a measure of "resourcefulness" or "precocious development" could easily be collected for 400 students. A major assumption underlying this study is, therefore, that *an objective, cognitive test is the fairest method of initially screening all students.*

Many cognitive tests have been justly criticized as being culturally biased. Some nonverbal tests, utilizing abstract items equally novel to all students, have been called "culture fair". Whether abilities measured by "culture fair" tests are in fact uninfluenced by such cultural variables as child-rearing practices is not established. It is nevertheless clear that the acquisition of vocabulary, for example, *is* subject to cultural influences; words you never hear or read you never learn. A second assumption made, therefore, is that *cultural differences are particularly evident on verbal or language tests*. The question of cultural bias would not be of concern in this study if all students in an inner-city school were assumed to be equally disadvantaged on verbal tests. Such an assumption is *not* made however.

Identification is presumably for the purpose of placement of the student in a gifted program. Some decisions about the program must be made clear since they affect the identification process. Firstly, what is the aim of the gifted program? Section 3822 states that a student is mentally gifted if he may be expected to "perform in school" at a level equivalent to that of the regularly identified gifted student. This seems to imply that the aim of gifted programs for disadvantaged students would be to produce achievement scores at the 98th percentile level. If the aim is high achievement, perhaps the selection should be based on an achievement measure. To dramatize the issue, consider a student with a nonverbal IQ of 135 but a very low reading level. Should he be placed in the program in preference to a student with a 120 IQ but who is one of the best readers in his grade?

This is a decision which schools must make. However, it is in the tradition of special education, whether for the retarded, the emotionally handicapped or the mentally gifted, to try to meet the needs of the highly deviant individual no matter what problems come along with his deviance from the norm. It is therefore assumed here that those of highest *ability* should be located for placement in a mentally gifted program. (Accelerated programs should exist for high achieving students.) However, this assumption has some implications for the nature of the gifted program. If it is to be a traditional lock-step program, placement in it of severe underachievers might damage the rate of progress of the whole program. In this case the decision is not at all easy. If individual help is not available for gifted but severly underachieving students, then placement of such a student in the gifted program is a risk. It will be assumed here that individual help, "adaptive treatment" to use the distinction made by Cronbach and Gleser (Cronbach 1957), is available even if the whole program is not individualized and, therefore, that *the aim of identification is to locate students of highest ability, regardless of their present achievement.*

But can "ability" be measured or are ability measures nothing but achievement measures? For the practical concerns of this paper it is assumed that a test which measures skills specifically taught in the schools can be called an [illegible] measure, whilst one which consists of items less subject to educational influ[illegible] be considered an ability measure. It is not assumed that any test measures the limit of a child, his genetic endowment or innate potential.

Lastly, it is assumed that the aim is to identify a full two percent of a grade level, the maximum permissible.

CHAPTER II

DESIGN OF THE STUDY

The sample

The eighth grade of an inner-city junior high school was selected for study. The school is approximately 87 percent Black and 10 percent Chicano. Since the eighth grade consisted of about 400 students, the problem under investigation was the location of the eight students of highest ability. (Eight is two percent of 400.)

The measures

Reliance on existing data for screening was rejected for several reasons. First, many records are incomplete or missing, particularly at the beginning of a school year. Secondly, records are often not uniform. Students from different school systems, for example, have taken different tests at different times. There is no satisfactory way to evaluate, for example, a CTMM score in fifth grade against a Lorge-Thorndike score in sixth grade. Finally, maturation rates might vary. Is the student who is average in the sixth grade necessarily average in the eighth grade or might he have "blossomed"? For these reasons it was decided that only current test scores should be used for screening.

Four measures were selected for concurrent administration:

1. The Standard Progressive Matrices (SPM) a nonverbal test which has frequently been used as a "culture fair" measure.

2. The California Test of Mental Maturity (CTMM) 1963 edition, Short Form, level 3. The CTMM is a conventional intelligence test widely used in schools.

3. The California Achievement Tests (CAT) 1970 edition, level 4, form A in mathematics and reading.

4. Teacher nominations. Although teacher nominations have been found to be inefficient and ineffective in locating mentally gifted students, this measure is still in frequent use in the schools.

It was originally planned that the SPM, CTMM, CAT and teacher nominations would all be investigated as means of identifying mentally gifted students. It was quickly realized, however, that none of these measures either discriminates sufficiently at the upper end of the score distribution or is reliable enough to be used to select just eight students.

Moreover, it was realized that any serious program for mentally gifted students should involve the administration of an individual intelligence test, such as the WISC, to each student. This being the case, the individual intelligence test would probably be the best criterion for identification.

However, under the assumptions made in Chapter I, a *full scale* score would be subject to cultural bias due to such verbal subtests as vocabularly and information. Consequently, it was decided that the WISC *performance* scale would be the criterion. Support for this choice is provided by recalling that a performance score on an individual test is the only objective criterion specified in section 3822. If a performance score is to be the criterion for some students it would seem consistent to make it the criterion for all students. Having decided upon the WISC performance scale IQ as a reasonable criterion, the problem becomes one of developing a procedure to select from the 400 those who should receive the WISC. (The WISC takes about an hour of a psychologist's time and so is too time-consuming and expensive to be given to all students.)

The proposed selection procedure for the WISC

In conformity with the assumption that culture fair, nonverbal measures should be used where possible, it seemed that the most desirable initial screening measure would be the SPM.

The SPM alone, however, was likely to lack sufficient differentiation in the top group of scores. Indeed, the SPM manual (Raven 1960) states that "the scale is intended to span the whole range of intellectual development, rather than to differentiate clearly between individuals." Consequently, it was decided to administer the more difficult level of the test, the Advanced Progressive Matrices (APM) to students who scored high on the SPM. This test is designed to "differentiate clearly between individual persons of even superior ability" (Raven 1965).

Furthermore, the use of a second test before final selection for the WISC has several important benefits. Since the number of students to be examined on the APM would be small enough to fit into one room, the administration of this second test could be made uniform for all students. Conducted, for example, in the library by a counselor, the possible influence of different times of day or classroom atmospheres, which could be cited as faults in the administration of the SPM, would be eliminated.

Thus, the proposed selection procedure involves the administration of the SPM in the various classrooms to all students, followed by the administration of the APM to those who scored high on the SPM. From these scores a group is selected to take the WISC.

This method would have to be rejected, however, if the SPM or APM showed serious deficiencies, such as lack of reasonable reliability or if the procedure did not efficiently and effectively select students who score highest on the WISC performance scale.

In order to validate the selection procedure, an effort must be made in this investigation to locate all those of high ability. Consequently, students scoring high on the CTMM, CAT and teacher nominations will receive the WISC in addition to those selected by the SPM-APM sequence.

Summary

The WISC performance scale was chosen as the criterion for the identification of mentally gifted students in the eighth grade of an inner-city junior high school.

A selection procedure was proposed for locating those students who are to take the WISC. This proposed selection procedure involves the administration of the SPM in the classrooms to all students followed by the administration of the APM to students who score high on the SPM. An examination of the relationship of these scores to the WISC scores will be made to see if this selection procedure can be made effective and efficient in selecting students who will score high on the WISC performance scale.

In this study, students will not only be selected for the WISC by the proposed selection procedure but also by the CTMM, the CAT and by teacher nominations.

CHAPTER III

PROCEDURES

Before any tests were administered, teacher nominations were collected by asking teachers to name "off-the-cuff" the eight students they considered to be of highest ability. Two methods were used to combine the nominations. First, the number of nominations received by each student was counted. Secondly, scores were assigned to nominations according to the rank the teacher gave the student in his list of eight nominees. A simple linear assignment of score to rank was used. The scores for each student were then added to yield a "teacher nomination score".

In Table 1, the composition and timing of the cognitive group tests is summarized.

The SPM was administered in the classrooms. Examples such as those reproduced in Appendix C were shown to each class before the test booklets were distributed. This introduction to the test was aimed at arousing interest and obtaining optimal effort on the test. A thirty-minute time limit was imposed. (The manual recommends the SPM be used as an untimed test, but this is not practical in a classroom situation.)

The APM was administered to small groups of students summoned to the school library. Set I was used as a practice set and items from it were discussed prior to administration of set II. The discussion given to each group was the same and is reproduced in Appendix C. Set II was given as a power test. The time taken by each student was noted.*

The CTMM short form was administered in the classroom. The only way it could be given in one fifty-minute class period was by omitting test 5. Test 5, a seven-minute test of mathematical word problems, was given to each class the day following the major part of the test. Thus, whilst the administration of the nonlanguage section was standard, the language section administration was not: test 7 the memory test, was answered after a shorter time than usual and test 5 was answered at a second sitting.

After the administration of the SPM and the CTMM to all 11 classes in the eighth grade and the collection of teacher nominations, economy of time required that the sample be reduced. In four classes, there had been no students scoring in the top two percent on any measure. These four classes were, therefore, dropped from the sample. The CAT tests administered to the remaining seven classes were the two reading tests (vocabulary and reading comprehension) and two of the mathematics tests (concepts and problems).

*This small group method of administering the APM is not in conformity with the method of administration recommended earlier. It was used because too many students were to receive the APM in this investigation to enable all to be tested at once, and because this method caused less disruption of classes.

TABLE 1

COMPOSITION AND TIMING OF THE SPM, APM CTMM AND THE CAT AS USED IN THIS STUDY

Test	Sub-Test	No. of Items	Working Time in Minutes
SPM		60	30
APM	Set I (practice set)	12	untimed
	Set II	36	untimed
CTMM non-lang.	1. Opposites	15	4
	2. Similarities	15	4
	3. Analogies	15	
	4. Numerical Values	15	[illegible]
CTMM lang.	5. Number Problems	10	7
	6. Verbal Comprehension (vocab.)	25	5
	7. Delayed Recall	25	6
CTMM	Total	120	37
CAT reading	Vocabulary	40	10
	Comprehension	45	40
	Reading Total	85	50
CAT math.	Concepts	35	10
	Problems	15	13
	Math Total	50	23

To select students for the WISC, the scores on the SPM, CTMM, CAT and on teacher nominations were rank ordered. The top eight students on each of these tests were scheduled for a WISC. This produced a group of 16 students. Five students were tested for other reasons: three because they had high APM scores and two for high scores on CAT reading or CAT mathematics.

The APM was given to students scoring above 48 on the SPM, to students selected for WISC testing and to a random sample of students. The random sample was selected by choosing every tenth name on an alphabetical list of the eighth grade.

Three classes were retested on the SPM after an interval of five weeks to enable calculation of a retest reliability measure.

After administration of the CTMM and the APM, a questionnaire asking for students' opinions about the test was given. (The questionnaire is reproduced in Appendix C.)

The administration of group tests is summarized in Table 2.

A rough measurement of SES level was made based on father's occupation as listed in guidance records and supplemented in some cases by information volunteered in interviews. The levels "high", "medium" and "low" are meant to refer to levels within this sample, not relative to the population at large. The judgments made are shown in Table 3. Information on SES level was collected only for the 22 students who took the WISC.

TABLE 2

GROUP TESTS ADMINISTERED

Test	Date	Sample	N
STANDARD PROGRESSIVE MATRICES (SPM)	October, 1971	total eighth grade (11 classes)	367
ADVANCED PROGRESSIVE MATRICES (APM)	November, December, 1971	a 10% random sample of Ss and 48 others	82
CALIFORNIA TEST OF MENTAL MATURITY, SHORT FORM (CTMM)	December, 1971	total eighth grade	308
CALIFORNIA ACHIEVEMENT TESTS:			
Reading: vocabulary and comprehension	January, 1972	seven* classes	179
Mathematics: Concepts and problems	January, 1972	seven* classes	207

*The four classes omitted were "low track" classes containing no students who had scored in the top 2% of the total sample on any previous measure.

TABLE 3

JUDGMENTS OF RELATIVE SES LEVELS FOR THIS SAMPLE, BASED ON FATHERS' OCCUPATION

High	Medium	Low
physician	shop specialist	laborer
supervisor	navy hospital	molder
administrative assistant	student worker phone clerk	machinist
public school or college administrator	lab. technician electronic technician	presser bodyman gardener
		unemployed

CHAPTER IV

RESULTS

The results section is divided into four parts; concerned with:

1. A general overview of the screening measures

2. The reliability and validity of the SPM

3. The APM and the WISC

4. The question of underachievement

A few notes on terminology may be helpful. Since the aim is to locate the top two percent which is about eight students, the top eight students on any measure are referred to as the "citerion group" for that measure. Thus, the "WISC performance criterion group" means the top eight students using the WISC performance test scores as the criterion. (In situations where there are tied scores, the criterion group may be more than eight.) To say that a measure "selected" a student means that the student was in the criterion group for that measure.

Reference will be made to "tracks". These are ability streams within the school. The single track I class is the highest ability class. There are three track II classes, four track III classes and two track IV classes. One class is a combined track II and III class.

The "accelerated Algebra program" is a class of about 35 students who take Algebra in the eighth grade rather than, as usual, in the ninth grade.

RESULTS - PART I: A GENERAL OVERVIEW OF THE SCREENING MEASURES

The seven eighth grade classes which took all screening tests consisted of approximately 200 students ranging in age at the beginning of the school year (September) from 151 months to 175 months. Mean age was 160 months (13 years, 4 months) and the standard deviation was 5 months. The sample was 50 percent male. Ethnic composition was 87 percent Black, 10 percent Chicano and three percent other.

The test most difficult to administer was the CTMM which involved constant timing and instructions. The SPM was very easy to administer and was very well received by students. Percent responses on the questionnaires filled in by students after taking the SPM and CTMM are shown in Table 4.

Twenty-two percent of students indicated on the post-test questionnaire that they liked the CTMM and 22 percent indicated that they did not like it. For the SPM, the corresponding figures were 37 percent (liked it) and eight percent (did not like it).

Whilst 11 percent thought they would have failed on the CTMM if it were graded, only one percent thought so on the SPM.

Distribution of obtained scores. Table 5 presents the results from the seven eighth grade classes which took all screening tests. Both the mathematics and reading achievement scores show a significant positive skew indicating a cluster of scores at the lower end of the scale and a "long tail" towards high achievement. The two nonverbal measures (the SPM and the CTMM nonlanguage test) show, by contrast, significant negative skew, indicating that on these scales fewer students received low scores and more scores clustered at the upper end of the distribution.

Table 6 shows the cut-off scores which define the criterion group (top two percent) on each test. Note that for both the total sample (Table 5) and the criterion group (Table 6) the percentile rank of the mean CTMM nonlanguage score is higher than that of the language score. The difference is considerable in the case of the criterion group: the mean raw score on the CTMM nonlanguage test is at the 92nd percentile point on national norms whereas the mean raw score on the language section is at the 69th percentile.

This observation conforms with the assumption that "cultural differences will be especially evident on language tests".

Illustration of the selection problem. It was suggested in Chapter II that none of the screening measures is reliable enough or shows sufficient discrimination of scores to be used to select just eight students. This statement is supported by the following examination of the actual results.

Table 6 shows the range of scores of the top eight students on each of the tests. A comparison of this range with the standard error of measurement, s_e, for each test

TABLE 4

PERCENT RESPONSES TO THE POST-TEST QUESTIONNAIRE

Question	Response on a Five Point Scale					Mean
	Strongly Negative			Strongly Positive		
	1	2	3	4	5	
1. How much did you like this test?						
after SPM	5	3	55	14	23	3.5
after CTMM	14	8	56	8	14	3.0
2. Were the instructions clear?						
after SPM	4	2	29	12	53	4.1
after CTMM	5	5	44	12	33	3.6
3. How well did you understand this test?						
after SPM	2	2	29	17	51	4.1
after CTMM	4	6	37	22	31	3.7
4. If the test were graded, what grade do you think you would get?						
after SPM	1	14	28	42	15	3.9
after CTMM	11	36	27	25	1	3.4
5. Did you have enough time?						
after SPM	1	14	28	42	15	3.6
after CTMM	11	36	27	25	1	2.7
6. Do you generally like taking tests?						
after SPM	10	32	37	19	2	2.7
after CTMM	10	29	42	18	1	2.7

TABLE 5

RESULTS ON THE CTMM, SPM AND CAT: TOTAL SAMPLE

Test		N	Mean Raw Score $\overline{X}$	Rank of $\overline{X}$ in Norms: Percentile	Stanine	SD	Skewness
CTMM	lang.	193	22.5	14	3	8.4	n.s.
	nonlang.	193	31.5	21	3	9.7	neg.*
	total	193	54.1	14	3	16.4	n.s.
SPM		208	39.9	37	4	7.1	neg.**
CAT	math.	207	18.1	19	3	7.4	pos.**
	reading	179	39.8	25	4	13.8	pos.*

**$p < .001$

*$p < .05$

TABLE 6

CRITERION GROUP SCORES ON THE CTMM, SPM AND CAT

Test		Cut-off Score	Range	s_e for Test	Mean Raw Score $\overline{X}_{cg}$	Rank of $\overline{X}_{cg}$ in Norms: Percentile	Stanine
CTMM	lang.	39	6	3.6	41	69	6
	nonlang.	48	5	3.7	50	92	8
	total	83	8	5.2	87	76	6
SPM		51	4	3.0	53	96	9
CAT	math.	32	7	2.9	35	73	6
	reading	67	6	3.7	69	83	7

indicates that much of the selection could be accounted for by errors of measurement. Many of the students scoring below the cut-off score might have true scores above it, and conversely, many scoring in the top eight might have true scores which would rank them much lower.

The problem can be illustrated by reference to the CAT mathematics scores (concepts and problems test combined) shown in Figure 1.

In Figure 1, the highest CAT mathematics scores are graphed, each score being represented by an x. If the obtained score is x, the probability that the true score is outside the confidence band given by $x \pm (1.96)\ s_e$ is less than .05. For the CAT mathematics test, the s_e is 2.9. Since (1.96) x (2.9) is approximately 6, a student with an obtained score of 32 can be said to have a true score in the interval 32 ± 6 with 95 percent confidence.

It would be unfair to make important decisions entirely on such uncertain data. This problem applies to each of the tests considered (in fact, it should be noted that the problem is even more severe for the other tests since the ratio s_e range is even larger). As an example of how reliable a test would need to be for such a low selection ratio as two percent, a test with a standard deviation of 10 raw score units would need a reliability of .99 in order that the 95% confidence interval be as small as four raw score units.

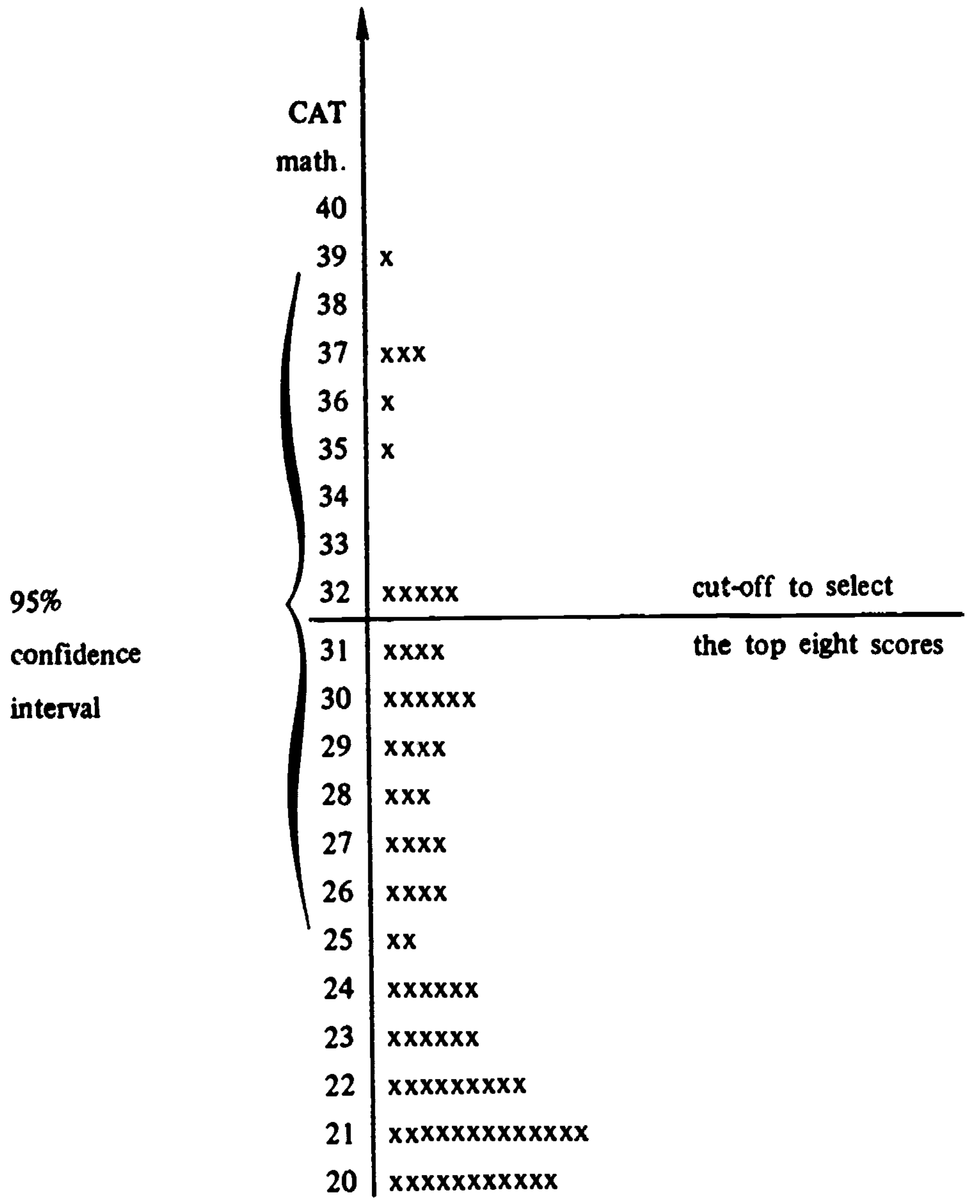


Figure 1

Illustration of the Selection Problem using CAT Mathematics Scores

RESULTS - PART II: RELIABILITY AND VALIDITY OF THE SPM

Reliability of SPM. Is the SPM as reliable as other available cognitive screening measures? Reliability coefficients are the most appropriate way of comparing the reliabilities of several tests.* Test-retest reliability is of primary interest in a selection problem, but where this coefficient is unavailable, the Kuder-Richardson formula 20 (KR20) coefficient will be used. (KR20 is the average of all possible split-half coefficients and as such is a measure of internal consistency.)

Three classes (N = 77) were retested on the SPM after an interval of five weeks. The Pearson correlation coefficient between the two tests was .86 ($p < .001$) a respectable test-retest reliability. Means and standard deviations for the two tests in this subsample are given in Table 7. A plot of scores from the two administrations (Figure 2) suggested that the relationship is not homoscedastic. The standard error is greater for lower scores than for higher scores, implying that higher scores are the more reliable.

The overall reliability of .86 and the indication that the reliability of higher scores is probably greater, make the SPM quite acceptable, with respect to reliability for use as an initial screening measure.

The other tests were not administered a second time, and the only way to compare their reliabilities is to consider reliability estimates from the manuals. These are KR20 measures. No corrections are available for SES or rural/urban differences between this sample and the norming populations. However, the reliability estimates can be corrected for differences in the variance of obtained scores. Column 5 of Table 8 presents the corrected reliabilities for the several tests.

In evaluating the corrected reliability samples it should be noted that KR20 estimates are generally higher than retest measures of reliability. Consequently, the SPM retest reliability coefficient of .86 compares satisfactorily with the CTMM (total) KR20 estimate of .89. The most reliable test is, not surprisingly, the longest test, CAT reading. There seems to be no reason to reject the SPM on the grounds of unreliability.

Validity of the SPM. In Table 9, intercorrelations among screening measures are shown. The correlations reported indicate that the SPM shows congruent validity. Correlations with other ability measures and with achievement measures are moderate, ranging from .42 to .60. It does not appear to "predict" achievement as well as the other ability measures, the CTMM language and nonlanguage tests. The CTMM language test, for example, correlates .65 and .82 with CAT mathematics and reading, respectively. However, this is not surprising when it is remembered that the CTMM language test contains a mathematics word problems test directly comparable to CAT mathematics problems and a vocabulary test (called "verbal comprehension") directly comparable to CAT reading vocabulary. (In this sample the two vocabulary measures, CTMM test 6 and CAT "verbal comprehension" correlated .81.) Thus, the CTMM language test "surreptitiously contains in itself most of the criterion it is claiming to predict" to use Cattell's words (Cattell 1963).

*Standard errors, being in raw score units, are not directly comparable between tests.

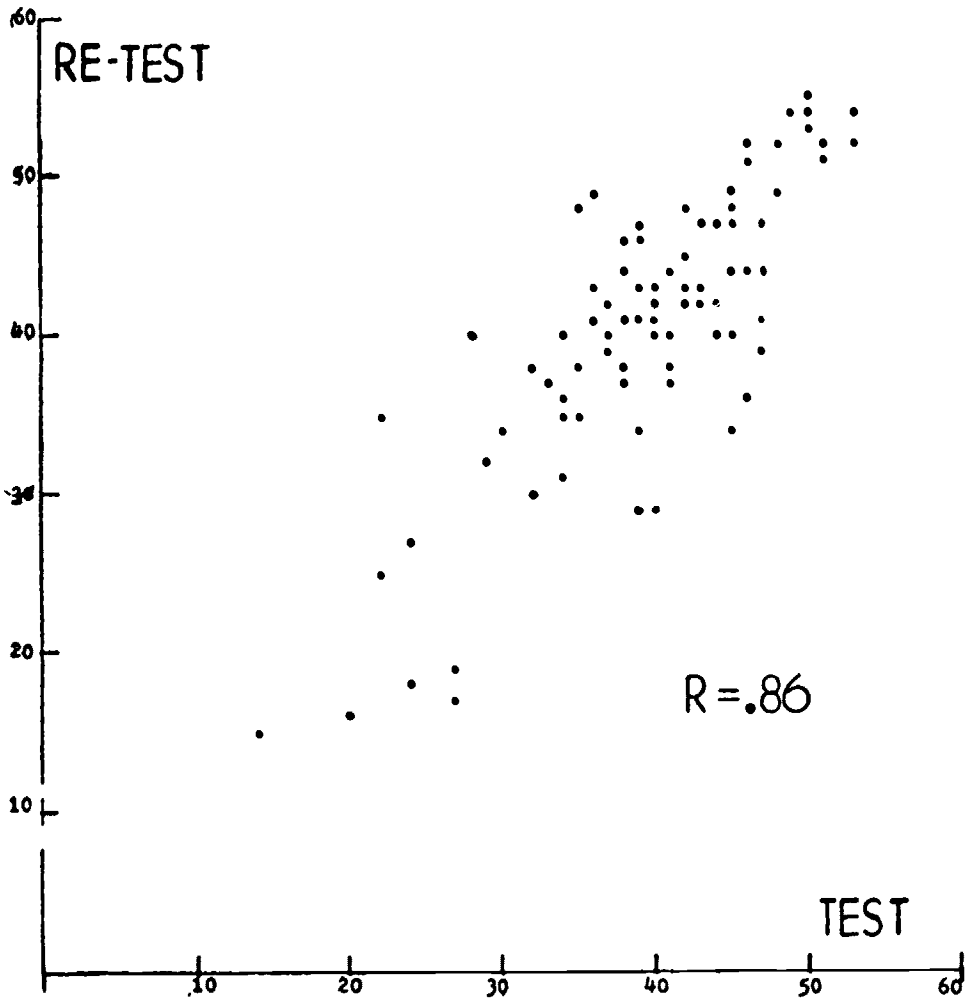


Figure 2

Graph of Test-Retest Scores for SPM

TABLE 7

TEST-RETEST RESULTS FOR THE SPM
N = 77

	Mean Score $\bar{X}$	SD
Test	38.1	7.9
Retest	40.4	9.9

Correlation = .86

TABLE 8

RELIABILITIES OF CTMM AND CAT CORRECTED FOR SAMPLE VARIANCE

	Test	Norming Population Data from the Test Manual r_{tt}	SD	SD this Sample	Reliability* Corrected for Sample Variance
CTMM	lang.	.90	11.3	8.4	.82
	nonlang.	.81	8.4	9.7	.84
	total	.91	17.8	16.4	.89
CAT	math.	.92	9.9	7.4	.85
	reading	.95	16.1	13.8	.92

*$r_{uu} = 1 - \frac{S_t^2 (1 - r_{tt})}{S_u^2}$ (Magnusson p. 75)

TABLE 9

INTERCORRELATIONS (SPEARMAN'S RHO) AMONG
THE SCREENING MEASURES
N ≃ 200

Measure	2	3	4	5	6
1. SPM	.42	.51	.60	.49	.26
2. CTMM language		.63	.65	.82	.42
3. CTMM nonlang.			.65	.60	.33
4. CAT math.				.70	.44
5. CAT reading					.38
6. Teacher nomination score					

NOTE: $p < .001$ for all figures

The SPM, on the contrary, contains not a word, not a number and yet shows predictive validity for achievement tests, particularly in mathematics.

A factor analysis of the tests should further elucidate the nature of the SPM. In Table 10, factor loadings after varimax rotation are shown for the SPM subtests, the CTMM subtests and the CAT tests. Loadings above .40 are shown for four factors which account for 65 percent of the variance. The first factor could be called numerical reasoning since both the CAT mathematics tests and tests 4 and 5 of the CTMM load on it. Tests 4 and 5 of the CTMM are reported in the CTMM manual to constitute a numerical reasoning factor, one of four statistically derived factors of the CTMM. It is on this numerical reasoning factor that the SPM loads most consistently. It also loads on factors III and IV which together seem to compose a factor which the CTMM manual calls logical reasoning, measured by the first three tests of the CTMM nonlanguage scale.

Thus, over the total range of scores the SPM appears to have measured an ability which might be called reasoning and which is related, in this sample, to mathematics achievement.

The correlations between tests and the factor analysis indicate agreement between the tests over the total range of scores. However, the main concern in using the SPM to screen for students of high ability is its validity at the upper end of the distribution. What kind of students score high on this measure? Do students who score very high simply possess a not very useful specific skill? How do the students who score highest on the SPM compare with students who score highest on other measures?

In order to answer these kinds of questions the top eight students on the SPM are compared with the top eight students on the CTMM, the CAT, teacher nominations and the WISC. (The full report of the WISC results is postponed to Part 4 of the results.)

Table 11 lists the students, represented by letters, in each criterion group and Table 12 summarizes how many "overlaps" there are between these groups. (One student occurring in the criterion group of each of two tests constitutes one "overlap" for each test.) On the basis of these two tables the SPM does not suffer in comparison with the other ability measure, the CTMM. It has more overlaps than the CTMM, and only one of the students selected by the SPM is not on any other list. Two of the students selected by the CTMM are not on any other list.

Six of the eight students in the SPM criterion group are in the CAT criterion group, that is they are among the highest achievers in their grade.

It has already been mentioned that the screening measures are not reliable enough to select the top eight students with any certainty. Nevertheless, these rough comparisons serve to indicate that the SPM does not lose validity at the extreme upper end of the distribution for this sample of students. Students who scored very high on the nonverbal, "culture fair" SPM usually showed other signs of high ability.

TABLE 10

ROTATED FACTOR LOADINGS FOR SPM, CTMM
AND CAT SUBTESTS
N = 118

	Variables	Factors I	II	III	IV
SPM	Set A			.58	.46
	Set B	.53			.49
	Set C	.42		.47	
	Set D	.41		.63	
	Set E	.69			
CTMM nonlang.	1. Opposites			.75	
	2. Similarities				.73
	3. Analogies		.45		.61
	4. Numerical Values	.67			
CTMM lang.	5. Number Problems	.82			
	6. Verbal Comprehension (vocab.)		.81		
	7. Delayed Recall		.70		
CAT	Reading Vocabulary		.74		
	Reading Comprehension	.45	.63		
	Math. Concepts	.65	.46		
	Math. Problems	.69			
PERCENT OF VARIANCE		40.9	10.4	7.3	6.3

TABLE 11

THE TOP EIGHT STUDENTS PLACED IN RANK ORDER BY EACH OF FIVE MEASURES

WISC Full Scale IQ	CTMM Raw Score Total	SPM Raw Score	CAT Reading + Math.	Teacher Nominations Score
A	A	H	E	E
B	I	E	F	A
C	L	T*	O	L
D*	V*	J	A	U
E	F	A	B	I
F	K*	F	H	R*
G*	B	O	C	X*
H	H	U	J	Q
	Q			

*Student selected by one measure only

NOTE: Brackets denote tied scores

TABLE 12

OVERLAPS* OF CRITERION GROUPS

Criterion	Total # of Overlaps	Overlaps With CTMM	SPM	CAT	Teacher Nominations
WISC	15	3	4	6	2
CTMM	15		3	4	4
SPM	16			6	3
CAT	18				2
Teacher Nominations	11				

*An "overlap" is a student selected by both of two measures. Thus, in Table 3, three students are selected by both the CTMM short form and the WISC; two students are selected by both the CAT total and the teacher nominations.

RESULTS - PART III: THE APM AND THE WISC

In this section, the APM results are reported for a random sample* and for those selected by high scores on the screening tests. The reliability of the APM and its correlation with the SPM are examined. The WISC results are then reported, and some characteristics of the gifted students (i.e., those in the top two percent using the WISC performance scale as the criterion) are noted.

The effectiveness of the APM in discriminating between high and low WISC scores is tested by discriminant analysis, and cut-off scores are selected for the proposed selection procedure.

The APM. The random sample which took the APM, consisted of 48 students ranging in age from 156 to 181 months. (Mean age was 160 months or 13 years, 7 months at the time of testing.) The mean scores of this random sample on the SPM and the CTMM were not significantly different from those of the total sample. The random sample thus appears to have been representative of the total sample.

The scores obtained by the random sample are reported in Table 13. Mean score on the APM was 12.5 with a standard deviation of 4.7. Reliability, for this sample, was .79 as measured by analysis of variance (Table 14).

If students who took the APM because they had obtained high scores on the SPM or some other test were included with the random sample, the reliability for this group of students was .83 (Table 15). Since the APM is designed for use in the upper ability range, it is this figure of .83 which will be used.

The standard error of estimate for this group was 2.3. Since the range of scores for the criterion group was only 5 (the top eight students scored between 22 and 27), the discrimination of scores at the upper end of the scale was little better than that of the SPM or the other screening tests, as can be seen by referring back to Table 6.

Table 16 shows correlations of the APM with other measures. As was the case with the SPM, the APM showed moderate congruent validity in this sample.

If the APM is to follow screening by the SPM, it is important to know how the two are related. In Figure 3, a scatter diagram shows the relationship for the combined random sample and selected students who took the APM. The correlation between the SPM and the APM was .69. The linear equation for the regression of the APM (Y) on the SPM (X) was:

$$Y = 0.46\,X - 3.7$$

*Since the APM (1962 edition) has not been widely used, and the manual is quite inadequate, the use of a random sample was desirable to establish some baseline data and to obtain an estimate of reliability unaffected by restriction of range.

TABLE 13

SCORES ON THE APM, SPM AND CTMM OBTAINED BY THE RANDOM SAMPLE

	Test	Mean	Range	SD
APM	Set I	8.3	2 to 12	2.4
	Set II	12.5	4 to 22	4.7
	Time on Set II (mins.)	34.1	14 to 71	14.2
SPM		38.4	12 to 52	8.6
CTMM	Total Raw Score	54.1	18 to 91	20.1

TABLE 14

ANALYSIS OF VARIANCE FOR APM SCORES: THE RANDOM SAMPLE

Source of Variation	Sum of Squares	df	MS	F
Between People	35.09	47	.75	18.30
Within People	359.86	1680	.21	
Questions	100.85	35	2.88	
Residual	259.00	1645	.15	

$$r_{tt} = \frac{.75 - .16}{.75} = .79$$

TABLE 15

ANALYSIS OF VARIANCE FOR ALL APM SCORES

Source of Variation	Sum of Squares	df	MS	F
Between People	87.06	95	.92	47.72
Within People	769.81	3360	.23	
Questions	257.40	35	7.35	
Residual	512.41	3325	.15	

$$r_{tt} = \frac{.92 - .15}{.92} = .83$$

TABLE 16

CORRELATIONS (SPEARMAN'S RHO) OF THE APM WITH OTHER MEASURES

Test		N	Correlation with APM
SPM		81	.69
CTMM	nonlanguage	66	.51
CTMM	language	58	.52
CAT	mathematics	45	.67
CAT	reading	45	.66

NOTE: $p < .001$ for all correlations

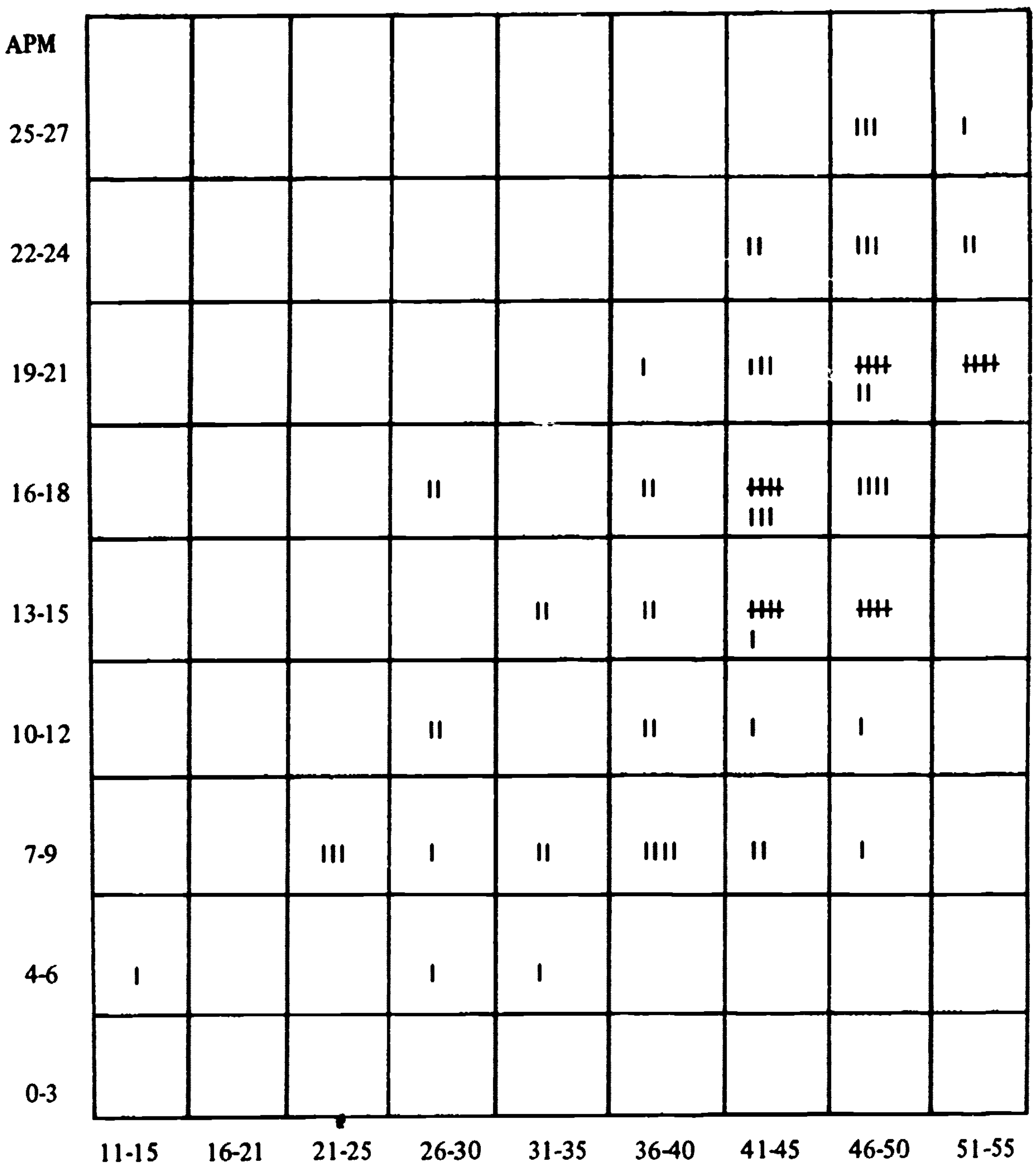


Figure 3

Scatter Diagram Showing Correlation
of SPM and APM

The standard error of estimate in predicting the APM score from the SPM was 4.0.

The WISC. Twenty-two students received a WISC; 13 male and nine female. A student was selected for the WISC if he had scored in the top two percent of the sample on one or more of the screening measures. The results, order by performance scores, are presented in Table 17. (Note that students are designated by letters which were assigned in the order of their full-scale WISC scores. Thus, student A had the highest full-scale score and B the second highest. B's performance scale score, however, ranked him 12th.)

In a disadvantaged population a mean measured IQ of 85 and a standard deviation of 15 is considered usual. This would imply that the 98th percentile point would be 116. The fact that the top eight students tested had full-scale scores at or above 115 suggested that the testing program probably picked out the students scoring in the uppermost two percent of the approximately 400 eighth grade students screened.

In the 22 WISC scores reported here, the performance scores were generally higher than the verbal scores, particularly among the higher total scores. Large discrepancies of 10 points or more (which is approximately twice the standard error of measurement for the WISC) occured five times in favor of performance scores (discrepancies of 28, 26, 24, 14 and 13 points) but only three times in favor of verbal scores (discrepancies of 19, 13 and 10 points). This conformed with the assumption that verbal scores would be depressed due to cultural differences.

The WISC performance criterion group. To select the top eight students on the performance scale required going down to a score of 114 at which point there was a tie so the group was expanded to include nine students, hereafter called the gifted students.

This group of nine students included seven of the eight students with the highest WISC *full scale* scores. The one student not included was student B.

Two students, A and G who will be called Albert and Gail, had WISC performance scale IQs above 130 and therefore qualified as gifted under the only objective criterion specified by section 3822 of the California State Code.

Student A. Albert was irrepressible. No matter what cognitive test was given he scored in the top two percent. Recognized as bright, he was nominated by his mathematics, science, English and music teachers. He is in track I classes and the accelerated Algebra program. He likes school and wants to go to college and enter a profession. His favorite subject is mathematics.

Gail presents quite a contrast to Albert. She represents a severly disadvantaged, extremely underachieving and yet definitely mentally gifted student.

Student G. Gail's high nonverbal ability was well established by four separate scores: the SPM, APM, CTMM nonlanguage and the WISC performance score. On the SPM,

TABLE 17

RESULTS ON THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN (WISC)

Student	Sex	Performance	Verbal	Full Scale
A	M	136	110	125
G	F	131	103	117
C	F	127	113	121
D	F	121	118	121
E	M	118	119	120
H	M	118	109	115
I	M	117	108	113
F	M	114	121	120
J	M	114	110	113
M	M	113	100	107
T	F	111	87	99
B	M	110	129	122
K	M	110	109	110
R	M	106	97	101
L	F	104	113	109
N	F	104	109	107
O	F	104	105	105
S	M	104	96	100
P	M	100	109	105
Q	M	97	110	104
U	F	97	100	99
V	F	94	104	99

she scored 50, missing th criterion group by just one point. On the APM, she scored 25, the second highest score in the sample. Her IQ score on the CTMM nonlanguage scale was the highest in the sample, 131, and the same as her WISC performance IQ.

Her WISC verbal IQ of 103 was based on only five subtests; the vocabulary subtest was invalidated since she declined to attempt any of the questions. Her CTMM language score yielded an IQ of 5. These items suggest severly depressed verbal achievement. She does not like to read, and this leads her to dislike school much of the time, especially classes in which she must read and write. Mathematics is her favorite subject, and she would like to be a mathematics teacher. On the CAT mathematics "problems" test she solved nine of the 15 word problems which ranked her tenth in this sample (the highest score was 13 out of the 15). The CAT mathematics "concepts" test, which involved a kind of mathematical vocabulary test and loaded on the verbal factor in the factor analysis, pulled her total score down. Her total CAT mathematics score was at the 50th percentile point on national norms which was well above the average for this sample. Her CAT reading score was at the 25th percentile point on national norms which was the rank of the average score in this sample.

Gail is in track II classes, not in the accelerated Algebra program and was not nominated by any teacher.

The other students in the WISC performance criterion group seemed to fall into two categories. There were the high scoring and recognized students (C, E, F and I) who, like Albert, received high scores on most measures and were among the top 16 on the teacher nomination scores. Then there were the high scoring and unrecognized students (D, H and J) who were not nominated by teachers although they also received high scores on most measures. These three students were not in track I classes or the accelerated Algebra program.

Table 18 shows how students, who took the WISC, scored on other measures. Scores on each test have been classified as 2, 1 or 0 according to whether the score was among the top eight scores (2), in the top 16 scores but not in the top eight (1), or below the cut-off score for the top 16 (0). The reason for the choice of rank 16 as a cut-off is a practical one. In California, the state reimburses a school district $40.00 for the cost of identifying a gifted student. Since an individual test costs about $20.00, the district can afford to test about twice as many students as are finally identified as gifted, i.e., an efficiency* of about 50 percent is financially acceptable. Consequently, if any one test is to be used for screening, the pertinent question is how many gifted students score in the top four percent on that measure, which in this case means in the top 16.

$$^{*}\text{Efficiency} = \frac{\text{Number of students found to be gifted}}{\text{Total number of students tested}} \times 100$$

TABLE 18

THE RANKS OF STUDENTS IN THE WISC SAMPLE ON OTHER MEASURES

Student	SPM	APM	CTMM Non-language	CTMM Language	CAT Math.	CAT Read.	Teacher Nomination Score
A	2	1	2	2	2	1	2
G	1	2	2*	0	0	0	0
C	1	2	2	0	1	2	1
D	1	1	0	1	1	2	0
E	2	1	0	2	2	2*	2*
H	2*	2	0	2	1	2	0
I	0	1	1	2	1	1	2
F	2	1	2	1	2	2	1
J	2	1	1	0	1	2	0
M	0	2*	0	0	NA	0	0
T	2	1	0	0	1	1	0
B	0	0	0	2	2*	0	0
K	0	0	2	0	0	0	0
R	0	0	0	0	0	NA	2
L	0	0	2	2	0	0	2
N	1	1	2	0	0	0	0
O	2	2	0	1	2	1	0
S	0	2	0	0	0	NA	0
P	0	0	0	1	2	0	1
Q	0	0	1	1	2	0	1
U	2	1	0	1	0	0	2
V	0	0	1	2*	0	2	0
Cut-off Score of Top 16	49	19	45	35	30	63	–

TABLE 18 (Continued)

NOTES: (1) "2" indicates the student was in the top eight students. "1" indicates the student was in the top 16 though not in the top eight. "0" indicates the student scored below the cut-off score for the top 16 students.

(2) Scores are listed in order of the WISC performance scores.

*Top scores on this measure

NA: Score not available

Table 18 supports very strongly the use of the SPM as a screening measure. If one screening test had to be used alone, the SPM would have been the best in this group. The use of any measure other than the SPM or the CTMM nonlanguage test would have overlooked Gail and use of the CTMM nonlanguage test would have overlooked three other students.

Table 18 also illustrates the observations already made that, with the exception of Gail, the gifted students were achieving highly. Excepting Gail, not one failed to be among the top 16 on the CAT mathematics test and the CAT reading test. Scores on the CTMM nonlanguage and language scales were less consistent but generally high.

Some noncognitive characteristics of the WISC performance criterion group. No generalizations about the characteristics of gifted students in inner-city schools can be drawn from examining one group of nine students. For example, the fact that seven out of nine students mentioned mathematics or science as their favorite subjects (Table 19) could be a chance effect, could be due to the particular teachers involved or could be a typical result.* Only a larger sample could allow positive inferences to be made. However, when asked why a subject was mentioned as a favorite, the reasons given concerned the content of the subject not the characteristics of the teacher. For example, "I like a challenge." "I'm learning something new." "I just love working problems." These attitudes could be inspired by good teaching, of course, but they could also reflect a preference of gifted students for subjects rich in structure and reasoning.

Very favorable comments were made by students about the accelerated Algebra program. Two of the students not in this program, D and H, commented that their mathematics classes were too easy. It is likely that, had more students been in the accelerated program, the number choosing mathematics as a favorite subject might have been even higher.

Career choices of this group were varied. Two students mentioned engineering in terms of the mechanics of automobiles or rockets. Otherwise, career choices mentioned were draftsman, doctor, nurse, teacher, commercial artist, secretary, air hostess, football player, musican and writer.

Three of the group, when asked to recall elementary school years, mentioned problems of being teased. They showed an awareness of possible causes with remarks such as, "I'm the sort of person who's easily hurt." "I was skipped a grade, so I was younger and smaller than the other guys." "I used to get mad real easy. For what? If I lost a game!" (laughed)

---

*Perhaps the only inference justified is a rejection of the hypothesis that, given a group of mentally gifted disadvantaged students, the majority will never prefer mathematics and science to other subjects. Here we have a group that does.

TABLE 19

SEX, SES, FAVORITE SUBJECT AND SOME SCHOOL CHARACTERISTICS OF THE GIFTED STUDENTS

Student	Sex	SES	Favorite Subject	GPA	Citizenship Grade	In Algebra Program	Track	Number of Teachers Who Nominated
A	M	L	Math.	B	1	Yes	I	3
G	F	L	Math.	C-	2	No	II	0
C	F	L	Math.	B+	1	Yes	I	2
D	F	M	English	A-	1-	No	II & III*	0
E	M	M	Math.	A	1-	Yes	I	5
H	M	L	Science	B	1	No	II	0
I	M	H	Math. and Science	A-	1-	No	II	3
F	M	L	Heritage	B+	2	Yes	I	2
J	M	M	Gym and Science	B-	1-	No	II	0

*Class combines track II and III

The proposed selection procedure. It had been hoped that the APM would "spread out" the scores at the top end of the scale enabling decisions about individuals to be made which were less subject to test errors. However, as noted above, the scores on the APM were as clustered as on the other screening tests and the standard error of measurement was as large in comparison with the criterion group range.

Did the APM, nevertheless, discriminate between those who scored high on the WISC and those who scored low?

To answer this question, a step-wise discriminant analysis was made between high and low WISC groups.

The WISC scores were divided into two groups. The high WISC group consisted of the nine students who scored 114 or higher on the WISC performance IQ scale (i.e., the WISC performance criterion group). The low WISC group consisted of nine students who scored below 114. (Students M, S and R had to be omitted from the analysis due to missing data.)

Table 20 shows the mean scores of the high and low WISC groups on the SPM, APM, CTMM language and nonlanguage tests and the four CAT tests.

Except for the mathematics concepts test, all high-WISC means were higher than low-WISC means, but only three reached the .05 significance level. The most significant difference in mean scores was on the APM. This lent support to the original supposition that the APM, the most difficult test, would show the most discrimination. The other two tests which showed significant differences between means were the SPM and mathematics problems. Thus, the students who scored highest on the WISC performance scale were most different from other students who took the WISC in their Matrices ability (SPM and APM) and their achievement level in solving mathematical word problems.

When the discriminant function, using the APM score, was applied to the groups, 15 out of 18 students were correctly classified. This was a promising result.

Exactly how should the SPM and APM scores be used to select students to take the WISC? In Figure 4, the SPM and APM scores are graphed for all students who took the WISC. The scores of the gifted students are indicated by stars. It can be readily seen that all the gifted students scored 21+ (21 or higher) on the APM *or* 52+ on the SPM.

In this sample, the following selection procedure would have been both efficient and effective:

1. Test all eighth grade students on the SPM.

2. Give APM to all students who scored 49+ on the SPM (N=21).

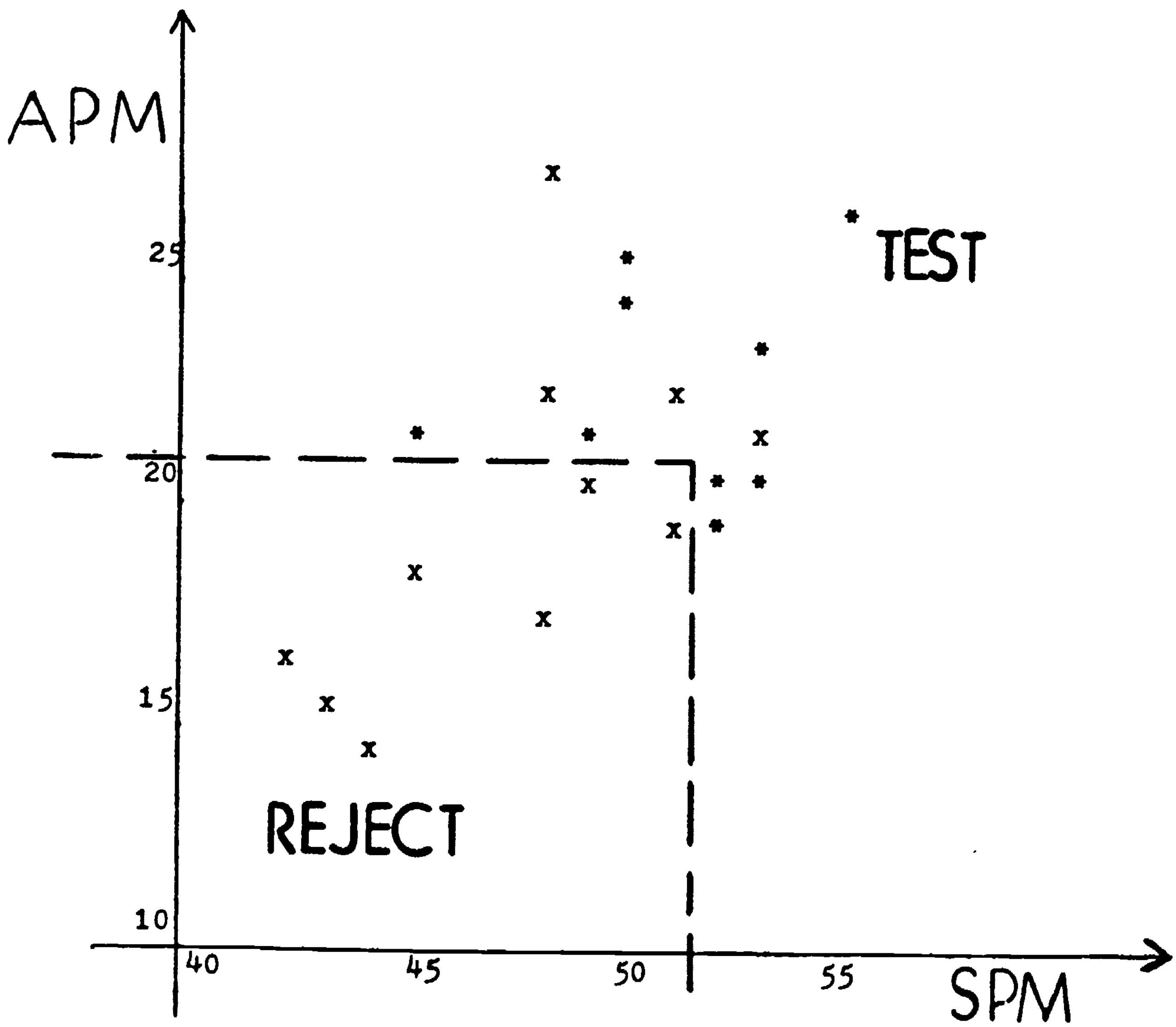
TABLE 20

MEAN SCORES OF THE HIGH-WISC AND LOW-WISC GROUPS ON OTHER MEASURES

Test		Mean Score of High-WISC Group	Mean Score of Low-WISC Group	F
SPM		51.0	45.0	5.12*
APM		22.1	17.7	13.56**
CTMM	Lang.	35.9	34.4	0.14
	Nonlang.	47.3	43.5	1.66
CAT Math	Concepts	21.0	22.1	0.47
	Problems	11.1	7.8	6.69*
CAT Reading	Vocabulary	32.7	29.2	3.19
	Comprehension	32.3	29.2	0.92

NOTE: $F_{.01,1,16} = 8.53$; $F_{.05,1,16} = 4.49$

*$p < .05$
**$p < .01$



*Gifted student
**Nongifted student

Figure 4

Graph of SPM and APM Scores for Students who Took the WISC

3. Select for the WISC all students who scored 21+ on the APM *or* who had scored 52+ on the SPM (N=14).

The number of students receiving a WISC would have been 14 of which eight were gifted (i.e., in the WISC performance criterion group). This is an efficiency of 56% and an effectiveness* of 89%.

The one gifted student overlooked by this process was student I who only scored 45 on the SPM and, therefore, would not have been selected for APM testing. However, if students nominated by two or more teachers had been added to the APM testing sample, this would have picked up student I. He scored 21 on the APM and would thus have been selected for WISC.

Based on this retrospective examination of WISC and matrices scores, a generalized selection procedure is shown in Figure 5. A discussion of the procedure is postponed to Chapter V. In Table 21, the scores made by individual students on the SPM, APM and WISC performance scale are reported.

$$\text{*Effectiveness of a process} = \frac{\text{\# of gifted students identified by process}}{\text{Total \# of gifted students}} \times 100$$

1. Administer SPM to all students as a 30 minute classroom test.

2. Administer APM as a power test to the top 6% on the SPM. Include also students strongly recommended by teachers.

3. Administer WISC to
   a) top 50% of APM sample
   b) students who were in the top 2% on the SPM.

4. The top students on the WISC performance scale are the gifted group.

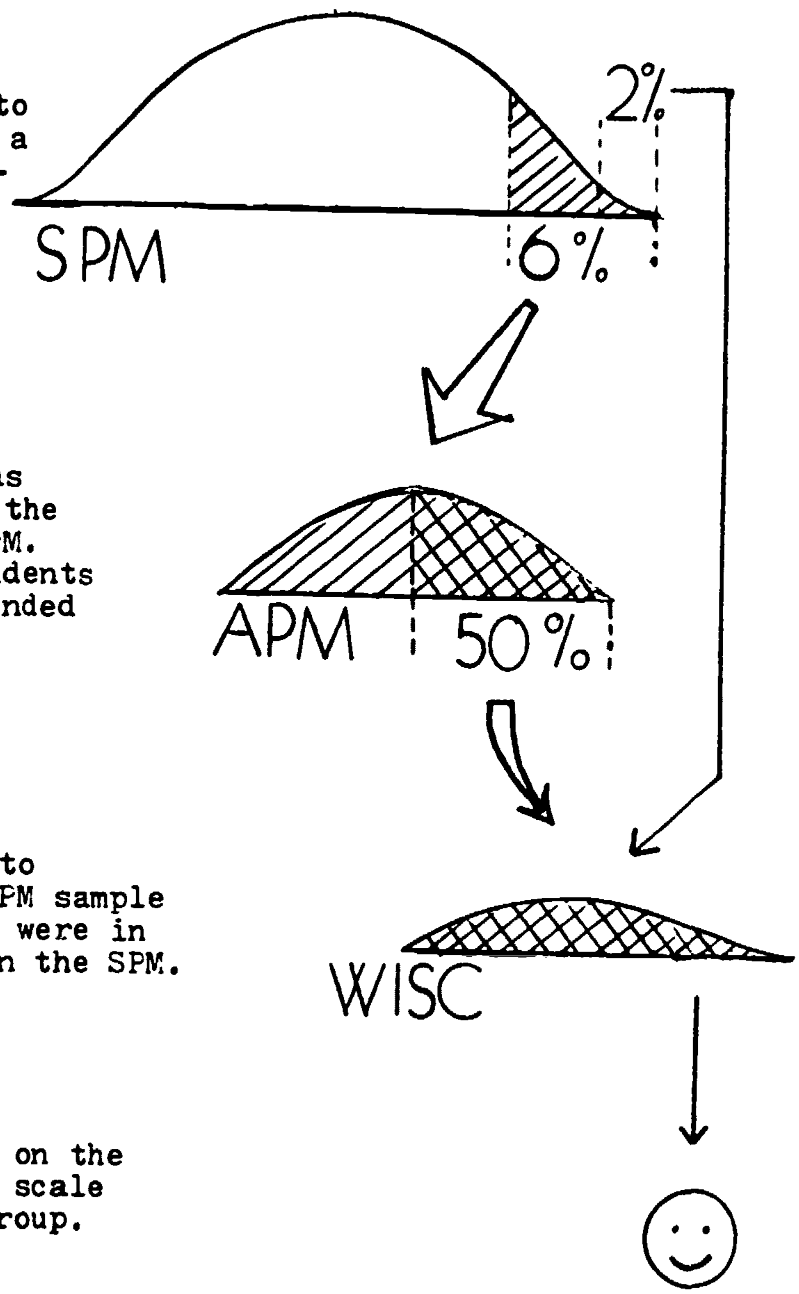


Figure 5

Recommended Procedure for Identifying Mentally Gifted Disadvantaged Students

TABLE 21

APM, APM AND WISC PERFORMANCE SCORES

Student	SPM	APM	WISC Performance Performance IQ
H	55	26	118*
J	53	23	114*
T	53	21	111
E	53	20	118*
A	52	20	136*
F	52	19	114*
O	51	22	104
U	51	19	97
-	50	27	NT[a]
G	50	25	131*
C	50	24	127*
-	50	24	NT
X	50	17	NT
-	50	14	NT
D	49	21	118*
-	49	21	NT
N	49	20	104
-	49	20	NT
-	49	16	NT
-	49	14	NT
-	49	13	NT
b			
M	48	27	113
S	48	22	104
L	48	17	104
I	45	21	117*
B	45	18	110
P	45	18	100
Q	44	14	97
V	43	15	94
R	42	16	106
K	28	17	110

a NT = not tested.

b Down to 49 on the SPM, all scores are reported. Below 49, only scores of students, who received a WISC, are listed.

*in WISC performance criterion group, i.e., gifted.

RESULTS - PART IV: THE QUESTION OF UNDERACHIEVEMENT

To be eligible for selection under the special criteria allowed by section 3822, title 5 of the California State Code, a student must not only be "culturally disadvantaged" but also "underachieving". Underachievement is generally understood as being indicated by a disparity between ability and achievement rank order scores. It is important to remember regression effects in such comparisons. If a group of students is selected for high percentile rank on one test, say test A, and these scores are compared with those on a second test, test B, the percentile scores on test B will tend to be lower than those on test A. This is because propitious errors will have accounted for some of the high percentile ranks on test A, the selection test. It is unlikely that propitious errors favor each student again on test B so, consequently, test B scores will generally be of lower percentile rank. As a consequence of this regression to the mean, gifted students (selected for high scores on an ability measure, test A) generally appear to be achieving, i.e., their percentile ranks on achievement measures (test B) are lower than those on the ability measure used in selection.

This effect can be avoided, however, by use of ability scores from a test other than the one on which the selection was based. In the present case students were selected as gifted on the basis of WISC performance scores. Four other "ability" measures can be used for comparison with achievement scores: the WISC verbal, CTMM nonlanguage and language, and the SPM. Table 22 shows the percentile ranks of the gifted students in the norming populations on these four ability measures and on the CAT achievement tests in reading and mathematics.

Using the WISC verbal scores as the ability measure, every student received a lower percentile rank on CAT mathematics. In fact, except for Albert and Gail, every student's percentile rank on the mathematics achievement tests was more than 10 points below his WISC verbal percentile rank. This pattern of underachievement was not repeated in the reading scores however. Comparing CAT reading percentiles with WISC verbal percentiles only two students (Gail and student F) showed an ability-achievement difference of more than 10 points, and six students received reading percentile ranks at or above their WISC verbal percentiles.

Using the "culture fair" ability measure, the SPM, all percentile ranks on both mathematics and reading achievement measures were lower, indicating underachievement. with the exception of student I.

To be rigorous, some measure of the reliability and significance of the differences would be needed, but such rigor is hardly warranted here since absolute faith is not placed in the ability measures, and the discrepancies are sufficient to meet the needs of the state criteria. Suffice it to say that the gifted students showed evidence of underachievement particularly in the field of mathematics and particularly if the SPM were used as the ability measure.

TABLE 22

PERCENTILE SCORES OF THE GIFTED STUDENTS ON ABILITY AND ACHIEVEMENT MEASURES

		Ability Measures			Achievement Measures	
Student	SPM	CTMM Nonlang.	CTMM Lang.	WISC Verbal	CAT Math.	CAT Reading
A	90	98	79	74	73	74
G	90	98	18	58	50	25
C	90	93	50	81	56	85
D	90	46	66	88	56	81
E	95	79	84	90	79	90
H	95	82	66	73	60	85
I	75	84	79	71	56	74
F	95	93	62	92	7[illegible]	79
J	95	76	54	74	60	81
average*	90	88	62	78	63	74

*This is the percentile rank corresponding to the average score, *not* the average of the percentile ranks. For the CTMM scores the average age of the group, 161 months, was used in finding the average percentile rank.

NOTE: These are percentile scores in the norming population for each test.

SUMMARY OF RESULTS

As suspected, none of the group measures showed sufficient discrimination of scores to be used alone to select the top two percent. Errors of measurement could account for most of the selection.

The SPM, the preferred screening measure since it is a "culture fair" test, was well received by students, easy to administer in the classroom as a thirty-minute test and showed a satisfactory test-retest reliability of .86. The mean score of 202 students was 39.9 with a standard deviation of 7.1. An apparent lack of homoscedasticity on the test-retest graph suggested that the standard error of measurement, 2.6 raw score units, should be less for higher scores.

Correlations of the SPM with the CAT and CTMM ranged from .42 to .60. These correlations, a factor analysis and an examination of scores at the top of each screening measure, suggested that the SPM measured, in this sample, an ability relevant to school achievement, particularly mathematics achievement.

The mean score on the APM, administered without a time limit to a random sample of 48 students with a mean age of 13 years 7 months was 12.5. The range was 4 to 27.

Analysis of variance of all APM scores (those of the random sample plus those of students selected to take the APM because of high scores on other measures) yielded a reliability estimate of .83.

Twenty-two students, selected for high scores on one or more of the screening measures, took the WISC. Two students received WISC performance scale scores over 130 thus qualifying them as mentally gifted under the only objective criterion specified by section 3822 of title 5 of the California State Code. One of these students was strong in all areas, scoring in the top two percent on every measure given. The other student, here called Gail, would have been entirely overlooked had the CAT, the CTMM total scores or teacher nominations been the screening measure. Her verbal scores were severely depressed (CTMM language IQ of 85) whilst her nonverbal scores were consistently outstanding. (e.g., CTMM nonlanguage IQ of 131; APM score of 25.)

As there was a tie for eighth place on the WISC performance IQ scale, the "gifted" group (defined as the top two percent using the WISC performance IQ as the criterion) consisted of nine students. With the exception of Gail these students were achieving in school at or near the top of their peer group on culture fair and conventional intelligence tests (the SPM and CTMM, respectively) and on reading and mathematics tests (the CAT). School grades were generally good though not exceptional. The grade point average was B or better for every student but Gail. Despite this generally high achievement as measured by both standardized tests and school grades, three of the gifted students in addition to Gail, were not nominated as possibly gifted by any teacher. Seven out of the nine gifted students mentioned mathematics or science as their favorite subject.

Using the WISC verbal score as an ability measure (to avoid regression effects since selection was based on the WISC performance score) there were strong indications that the gifted students were "underachieving" in mathematics but generally not in reading. Use of nonverbal ability measures generally showed underachievement in both mathematics and reading however.

The proposed selection procedure (the SPM followed by the APM) could have selected 14 students for the WISC eight of which were in the gifted group, indicating that it is an efficient and effective selection procedure. Although the APM did not "spread out" the high scores any better than did the other screening measures, discriminant analysis of high and low WISC scores showed it to be the most discriminating test. Students who scored high on the WISC differed most from students who scored low on the WISC in their scores on the APM, the CAT mathematics word problems test and the SPM, in that order.

CHAPTER V

DISCUSSION OF RESULTS

In this chapter, the assumptions made in the design of the investigation are examined in the light of the results obtained. The recommended selection procedure is then discussed and some comments made on the contrasting problems of special and normal criteria.

The assumptions in retrospect

The first assumption was that *an objective cognitive test for all students is the fairest initial screening measure*. This assumption was initially justified on the basis that if the criterion is to be performance on tests (as section 3822 of the California State Code would seem to indicate) the selection procedure should be based on test performance. However, the assumption can also be examined by consideration of the alternatives. One alternative to an objective cognitive test for all students is to ask for teacher nominations. Often these nominations are not used to make final decisions but as a starting point. Thus, the folders of the students nominated are then examined, or the students nominated are given tests or a committee meets to discuss them. In the present investigation, four of the nine gifted students were not mentioned by any teacher and thus would not have even entered into consideration had teacher nominations been used as a starting point.

Another tempting alternative to testing everyone is to test just the top, track I class. Surely the top eight students are to be found among the ones considered to be the top 35. Yet, five out of the nine gifted students were not in the top track. They were, however, in track II classes.

Nor did grades distinguish the mentally gifted students. Many other students had as high or higher grade point averages.

These findings conform with studies of the identification of gifted students in other populations. Pegnato (1958) for example, studying a high SES junior high school found teacher nominations both inefficient (many students were nominated who were not gifted) and ineffective (many students who were gifted were not nominated).

Bearing in mind these results and also the desirability of using a selection procedure which is above reproach, it seems reasonable to assert that in attempting to locate students who score high on an individual test, the screening of all students by an objective cognitive test is the best initial step.

A second assumption made was that *cultural differences will be particularly evident on verbal or language tests*. This assumption seemed to be well supported for students in the upper ability range. Table 3, for example. showed the criterion group mean scores on several measures. The highest percentile rank of these means in the norms was on the nonverbal, "culture fair" SPM (96 percentile) followed by the CTMM nonlanguage (92 percentile). Thus, the ability of these inner-city junior high school students relative to the general population was least depressed on nonverbal measures.

The pattern of WISC scores supported the same conclusion for students in the upper ability range. Differences of 10 points or more between WISC verbal and performance scores occured more often in favor of performance scores than in favor of verbal scores and the discrepancies in favor of performance scores were larger.

The third assumption was that *the aim of identification is to locate students of highest ability, regardless of their present achievement*. As it turned out, most students who scored high on the culture fair or conventional measures of ability were achieving highly relative to their peer group. For example, six out of the top eight on the SPM were in the top eight on the CAT total. But the assumption became pertinent in the case of one student, Gail, who was decidely outstanding on all nonverbal measures but to whom reading is a chore and vocabulary a mystery. Gail, with a WISC performance score of 131 but a CTMM language IQ score of 85 (to pick out two contrasting scores) is definitely an underachieving but mentally gifted student. Any screening process which overlooked her or students like her should be subject to severe criticism on that account.

The recommended selection procedure

If the WISC sample had been selected only on the basis of the proposed screening procedure, which involves only the matrices tests, it might have been biased and conclusions could only have been considered applicable to other samples selected in the same way. The question would still remain: were there others who would have also scored high on the WISC but who were overlooked by the matrices screening procedure?

However, 18 out of the 21 students who received the WISC were "selected" by measures other than the matrices tests. That is to say, if the matrices had not been used at all and students selected for the WISC by only the CTMM total, the CAT total and teacher nominations, 18 of the 22 students would have been selected. Consequently, it is reasonable to claim that the WISC sample was representative of the highest scoring students on several non-matrices measures.

If it can now be shown, in retrospect, that applying only the recommended selection procedure would have been reasonably effective in selecting the gifted group, this will be a strong indication of the validity of the selection procedure.

Of the 367 eighth grade students who took the SPM*, 21 (approximately six percent) scored 49 or over. These students took the APM. Selecting from this group on the criterion of *either* an SPM score of 52+ (six students) *or* an APM score of 21+ (10 students) yielded a combined group of 14 students which contained eight of the nine students in the WISC performance criterion group. (The one student overlooked was student I.)

Thus, the matrices tests, although requiring no school-learned skills such as reading or arithmetic, were nevertheless as good as conventional measures at picking up outstanding students. Moverover, a student whose reading and verbal skills were deficient was not overlooked as would have been the case with conventional measures.

The efficiency and effectiveness** of the proposed selection procedure was 89 percent and 56 percent, respectively, in this study These figures are remarkably high in comparison with the study by Pegnato (1958). Table 23 reports the highest efficiency and highest effectiveness ratios he found.

There are so many differences between this study and Pegnato's, however, that comparisons are really not in order. The data are quoted here because Pegnato's study is one of the most thorough studies conducted on the location of gifted students, and from it the outlook had seemed rather bleak. No screening measure he studied (the measures were teacher judgment, honor roll membership, creativity, student council membership, mathematics achievement, group IQ scores, and group achievement tests) which had an effectiveness above 70 percent, had an efficiency above 23 percent. The highest effectiveness ratio he found was 92 percent, but to get this it was necessary to test about five students for every one found to be gifted, a very uneconomical procedure.

In the present study not enough individual tests were given to allow very strong conclusions to be drawn, but the results suggest that a procedure involving the administration of a test at two levels can be satisfactorily efficient and effective in selecting students for individual testing.

In Appendix D, an attempted cross-validation of the procedure is described.

Just using the SPM alone, rather than as part of a two stage selection procedure, could have selected 13 students for testing of whom seven were in the gifted group, an effectiveness of 77 percent and an efficiency of 54 percent. This raises the question as to whether the addition of the APM was necessary or justified. The problem which arises in the use of the SPM alone is the rapid clustering of scores as the scale is descended. To this lack of differentiation must be added the problem of test errors. The SPM alone would not be a poor screening instrument, but the addition of the APM permits an increase in efficiency and effectiveness and a *great* increase in the justifiability and demonstrable

---

*This was all eighth grade students present on the testing day. Mean score was 37.8 with a standard deviation of 8.7.

**Effectiveness is the percent of gifted students identified out of the total number of gifted students. Efficiency is the percent of gifted students identified out of the number selected by the screening procedure.

TABLE 23

EFFICIENCY AND EFFECTIVENESS OF THE RECOMMENDED SELECTION PROCEDURE (THE MATRICES) COMPARED WITH SELECTED RESULTS FROM PEGNATO'S STUDY

Study	Screening Procedure	Effectiveness	Efficiency
	Group Intelligence tests		
Pegnato's	cut-off 115	92	19
	cut-off 130	22	56
Present	Matrices	89	56

fairness of the testing procedure. This is because use of the APM as recommended means there is one uniform test administration and students get two chances to score above a criterion score, once under timed conditions (the SPM) and once without any time pressure (the APM).

Was there any screening measure other than the matrices which showed equal promise? In this regard it must first be remarked that any one of the objective cognitive tests used in this study was more efficient and effective in selecting gifted students than were teacher nominations. However, every test but the CTMM nonlanguage scale would have overlooked Gail who, with a WISC performance IQ of 131, was undoubtedly gifted as defined by the special criteria. (The CAT mathematics word problems test would have selected her for testing, but it is a brief 15 item test and is usually combined with the mathematics concepts test. On the combined score Gail was not among the top 16 students.) The use of the CTMM nonlanguage test to select 16 students for testing would have overlooked gifted students D, E and H. Consequently, on the basis of this investigation, the matrices screening procedure appears to be the most preferrable.

Contrasting problems of special and normal criteria

The situation in selecting students under the special criteria is rather different from the normal selection problem for gifted programs.

Normally, any student scoring above a criterion score, the 98th percentile point on an approved test, can be placed in a gifted program and the state reimburses the school district. This means the number to be placed in a gifted program is flexible, depending only on the number of students who can be found with scores above the criterion. This number varies significantly with the socioeconomic status (SES) of the school population. In the high SES school studied by Pegnato, for example, (Pegnato 1958) 6.5 percent of the school population had Stanford Binet IQs over 136, i.e., at or above the 99th percentile point for the general population. Such a school could support a gifted program enrolling a large number of students while a less favored school would scarcely have enough students qualifying as gifted to make a special program economically feasible.

In selecting under the special criteria, however, the number to be identified is more or less predetermined, being limited to two percent of the total disadvantaged population in the school. There is no specified criterion above which a student must score. Consequently, there is not the same incentive to search for all the most gifted students. No matter what measure is used for identification a school can usually designate two percent as gifted under the special criteria.

The fact that this makes "identification" easier should not lead to less effort being put into the identification process. Decisions based on unreliable test scores, or on subjective judgments that are little more than personal "hunches", might be doing great injustice to the child who is omitted from a program in which he should be placed. Unsupportable decisions deserve to be challenged and in view of current widespread dissatisfaction with testing in general, it is possible that they will be.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

A procedure is recommended for the identification of mentally gifted disadvantaged students.

The procedure involves the screening of all students on the SPM, a culture fair test which is easy to administer in classrooms and well received by students. The SPM is to be followed by the APM, a higher level of matrices test. Selected students are then given an individual test, the WISC, and the performance scale IQ is used as the criterion for identification.

No serious defects were found in this recommended procedure. The SPM appeared to be both reliable and valid as an ability measure. The APM permitted discrimination to be made between those likely to score high and those likely to score low on the WISC. Differentiation of scores on the WISC performance scale was satisfactory and better than on any of the group tests. Test scores on conventional ability and achievement measures showed the students identified as gifted to be of generally outstanding ability in their peer group.

The assumptions made in the design of the investigation seemed justified or necessary. Any objective cognitive test was preferable to teacher nominations in both efficiency and effectiveness. Verbal ability scores of high ability students were depressed relative to their nonverbal scores. The assumption that the aim of identification was to locate students of the highest ability regardless of present achievement became necessary in considering one student. Since she met the only objective criterion specified by the State Code (a performance scale score above 130) she should be considered gifted despite very low present achievement.

The proposed identification procedure appears to be fair and efficient. It would have worked well in this sample.

Need for further research. The use of the WISC performance scale as the criterion is defensible in that an individual test is the traditional method of identifying gifted students (because it is considered more reliable and valid than group measures) and a performance IQ is specified as a criterion for disadvantaged students by the California State Code. Ultimately, however, this criterion needs to be validated against performance in gifted programs of various types.

Further tests of the matrices procedure recommended for selecting students for WISC testing are also needed, especially ones involving the administration of a larger number of individual tests to check that highly able students are not being overlooked.

Summary of recommendations

1. The WISC performance scale IQ score is recommended as the criterion for the identification of mentally gifted disadvantaged students.

2. Reliance should *not* be placed on teacher nominations or noncurrent test data to screen students.

3. A cognitive test given to all students is recommended as a fair and objective screening procedure. In a disadvantaged population a culture fair or nonverbal test is to be preferred.

4. The following procedure, using two levels of the Progressive Matrices tests is recommended to select students to take the WISC:

a. Administer the SPM as a 30 minute test to all students in the grade. This can be done in the classrooms by teachers.

b. Administer the APM to students who scored in the top six percent of the grade on the SPM. The APM should be given at the same time and place for all students, i.e., in one uniform administration. (NOTE: Students strongly recommended by teachers can also be included in the APM testing.)

c. Select for the WISC students who scored in the top two percent of the grade on the SPM *or* in the top half of the APM sample.

(A practical, step by step guide to this procedure is given in Appendix B.)

CHAPTER VII

BIBLIOGRAPHY

Anastasi, A. *Psychological Testing.* New York: MacMillan Co., 1961.

Cattell, R.B. "Fluid and Crystallized Intelligence". *Journal of Educational Psychology.* 54 (1), 1963.

Pegnato, C. *An Evaluation of Various Initial Methods of Selecting Intellectually Gifted Children at the Junior High School Level.* Doctoral Thesis, Pennsylvania State College, 1958.

Raven, J.C. *Advanced Progressive Matrices.* London: H.K. Lewis and Co. Ltd., 1965.

Raven, J.C. *Guide to the Standard Progressive Matrices.* London: H.K. Lewis and Co. Ltd., 1960.

APPENDIX A

SECTIONS 3821 AND 3822 OF THE CALIFORNIA ADMINISTRATIVE CODE, TITLE V

3821. Evidence to Be Studied, Except for Culturally Disadvantaged, Underachieving Pupils. Except for culturally disadvantaged, underachieving pupils described in Section 3822, among the items of evidence studied by the committee shall be evidence described in either (a), (b), or (c):

(a) A score at or above the 98th percentile on a full scale individual intelligence test approved by the Superintendent of Public Instruction and administered to the pupil by a person qualified to administer individual intelligence tests. The norm to be used for the score is the norm for children of the same age as the pupil tested.

(b) For a pupil in grades seven through twelve a score at or above the 98th percentile *in each of two tests chosen from a list of tests* approved for the purpose by the State Board of Education and administered by a qualified person to the pupil while he was enrolled in grade 7 or above and *within 24 months* of the date of identification. The two required tests are:

(1) A standardized full-scale group test of mental ability.

(2) A standardized test of one of the following:

(A) Reading achievement
(B) Arithmetic achievement
(C) Science achievement
(D) Social science achievement

(c) The judgments of teachers, psychologists, and administrators and supervisors who are familiar with the demonstrated ability or potential of the minors. In any given district not more than five (5) percent of the pupils identified under Section 3821 as mentally gifted minors shall be identified on such judgments alone.

3822. Separate Criteria for Identifying Culturally Disadvantaged, Underachieving, Mentally Gifted Minors. A culturally disadvantaged, underachieving, mentally gifted minor shall be identified by the committee described in Section 3821 as follows:

(a) As "culturally disadvantaged" - through the committee's study of all available and pertinent evidence of a child's language, cultural, economic or environmental handicaps that have in the past and may in the future interfere with his success in school, restrict the development of intellectual and creative ability, and prevent full development of his potential. The report of the committee shall specify the disadvantage or disadvantages to which the pupil is subject. "Cultural disadvantage" shall be distinguished from cultural difference; it implies a lack of cultural background rather than a comparison among cultures.

(b) As "underachieving scholastically" - by comparing the pupil's general intellectual capacity with this achievement on the basis of all pertinent evidence related to cultural disadvantage. Consideration shall be given to each of the following:

(1) The judgment of the committee, all concurring, that the pupil could achieve at the upper two percent level were it not for his cultural disadvantage.

(2) Test scores revealing discrepancies between general intellectual ability and achievement.

(3) All pertinent school records.

(c) As "mentally gifted" - on the basis of the judgment of the committee that he may be expected within a reasonable time and with appropriate curricular modifications to perform in school at a level equivalent to that of the mentally gifted minors identified pursuant to Section 3821. This judgment shall be based upon one or more of the following:

(1) Precocious development and maturation in the preschool or primary period, or outstanding scholastic accomplishment at any point in his school career.

(2) Unusual resourcefulness in coping with responsibilities, opportunities, deprivations, problems, frustrations, obstacles, lack of structure and direction, or overly structured settings.

(3) Outstanding achievements, skills, or creative products.

(4) Scores at or above the 98th percentile on nonverbal (performance) scores of individual intelligence tests approved by the Superintendent of Public Instruction.

(d) In no case shall the minors identified pursuant to this section exceed two percent of the culturally disadvantaged pupils (as defined in Education Code Section 6452) within the school district.

APPENDIX B

PRACTICAL GUIDELINES FOR COUNSELORS ADMINISTERING THE RECOMMENDED PROCEDURE

1. Purchase or otherwise obtain the Standard Progressive Matrices (SPM). Cost is $2.00 per test. The number purchased should be equal to the largest class size (plus a few extra in case of wear and tear or defacement of course.)

2. Purchase or otherwise obtain the Advanced Progressive Matrices (APM). Cost is $1.50 for set I, the practice set and $3.50 for set II, the test. The number purchased should be equal to about seven percent of the number to be screened. (For example, if the eighth grade is being screened and there are 400 students in the eighth grade, purchase 28 APM's since 28 is seven percent of 400.)

3. Have answer sheets prepared. The number prepared for the SPM should equal the number of students in the grade being screened. The number prepared for the APM should equal the number of APM's purchased.

4. Select classes in which the screening tests will be given. The simplest arrangement is to select two or three teachers who between them teach the whole grade. For example, if everyone in the grade takes science, the test could be administered in science classes. Only two or three teachers would be involved in the testing so that test administration would be fairly uniform. However, it would mean that the test was administered to students at various times of the day. Performance might not be too good by the last period of the day when students are tired or restless.

Another approach would be to have the test given in, say, second period classes. This however, would mean involving many more teachers.

Note that the use of the APM will give students a second chance so that small differences in performance due, for example, to being tired at the end of the day are not critical.

5. When the tests arrive hold a meeting with the teachers involved in the administration who will be administering the SPM. Explain the testing program and the SPM and agree on a standardized testing procedure. This procedure could be simply the instructions given on page 9 on the SPM manual (Raven 1960). However, to generate enthusiasm it might be decided to use more sample questions, such as those reproduced in Appendix C of this report. These could be drawn on the blackboard in each classroom, reproduced on large poster boards by the art department or made into transparencies for use with an overhead projector. Items from set I of the APM could easily be copied for transparencies. It is recommended that only simple items, such as the first six of set I, be used since the purpose is to generate interest and confidence, not to teach students how to solve the items.

Whatever procedure is chosen, it should be uniform for all classrooms. Accurate timing of 30 minutes and care that no test booklet is lost should be stressed.

Select testing days and arrange who will get the tests and answer sheets from whom, etc.

Warn teachers that many students, particularly in the lower ability classes, will finish before the time limit. They may wish to make some provision for this.

Instruct teachers to turn in after testing:

(a) the answer sheets

(b) a list of all students enrolled in the class as of the testing date. Students who missed taking the test, due to absence or some other eventuality, should be identified on this list

6. When tests and answer sheets are turned in, check that no test booklets are missing. Have answer sheets screened and graded.

Screening of answer sheets. A stencil made to fit over the answer sheet greatly speeds screening. In checking answers start with E12, the most difficult item, and work backwards through the test. A paper can be set aside when 15 items have been marked wrong since a score of 45 out of 60 is almost certain to be below the top six percent of scores.

Grading of answer sheets. Score completely each paper in which less than 15 items are missed. Record the total score (number right) on the class lists provided by the teachers.

7. Arrange for the testing of those who missed the initial classroom administration of the SPM. Screen and grade these "pick-ups" and record on the class lists.

8. From the class lists derive a rank order listing of the top students on the SPM. Test the top six percent with the APM.

APM testing. The APM should be given to all selected students at the same time and supervised by a counselor. (For example, students can be summoned to the library for the APM.) Allow about 70 minutes total time.

(a) Pass out set I and answer sheets. Have students answer set I.

(b) When all students have finished, read a discussion of set I items. (A sample discussion is reproduced in Appendix C. Since it is necessary to point to items, transparencies for use on an overhead projector would be useful.)

(c) Pass out set II. Allow students to take as long as necessary to attempt all items.

9. Grade set II of each APM paper and record the results on the rank order list of SPM scores.

10. Determine the median score on the APM. Schedule for a WISC any student who was in the top half of the APM sample or who was in the top two percent on the SPM.

If time and money permit, schedule also those students nominated as possibly gifted by three or more teachers.

11. When all WISC results are available, select as the "special criteria mentally gifted" group the students with the highest WISC performance scores.

APPENDIX C

SAMPLE MATERIALS

This appendix contains the following three samples:

1. Questionnaire asking student opinions on tests.
2. The explanations, given orally, after students had attempted Set I of the APM.
3. Examples for introducing the SPM (X1, X2, etc.).

NAME ____________________

TEST ____________________

DATE __________ PERIOD __________

1. How much did you like this test?

/	/	/	/	/
not at all		OK		very much

2. Were the instructions clear to you? Did you always know what to do?

/	/	/	/	/
instructions were hard. I didn't know what was wanted.		OK		instructions were clear. I always knew what to do.

3. How well did you understand this test?

/	/	/	/	/
not at all		OK		very well

4. If the test were graded, what grade do you think you would get?

/	/	/	/	/
F	D	C	B	A

5. Did you have enough time?

/	/	/	/	/
needed a lot more time	needed a little more time	had just enough	had time to finish easily	too much time was given

6. Do you generally like taking tests?

/	/	/	/	/
I hate tests	no, not really	OK	yes	I love tests

THE EXPLANATIONS, GIVEN ORALLY, AFTER STUDENTS HAD ATTEMPTED SET I

Now you've all finished trying set I; we'll take a look at some of the questions. I'm going to read the explanations to you: this is to make sure everyone gets told the same things.

Let's look at No. 5. If we look at the background first and read across the rows (pointing) we see that the background doesn't change. In the top row it's all white, the second row it's half shaded, in the third row it's all horizontally shaded. So we know the missing piece must have a horizontally shaded background. Now look at the foreground. That changes with columns. There is nothing added on top of the background in the first column (pointing), in the second column there is a half vertically-shaded diamond, and in the last column there is a complete, vertically-shaded diamond. So the missing piece must have a complete vertically-shaded diamond as well as a horizontally shaded background. The correct answer, then is 2.

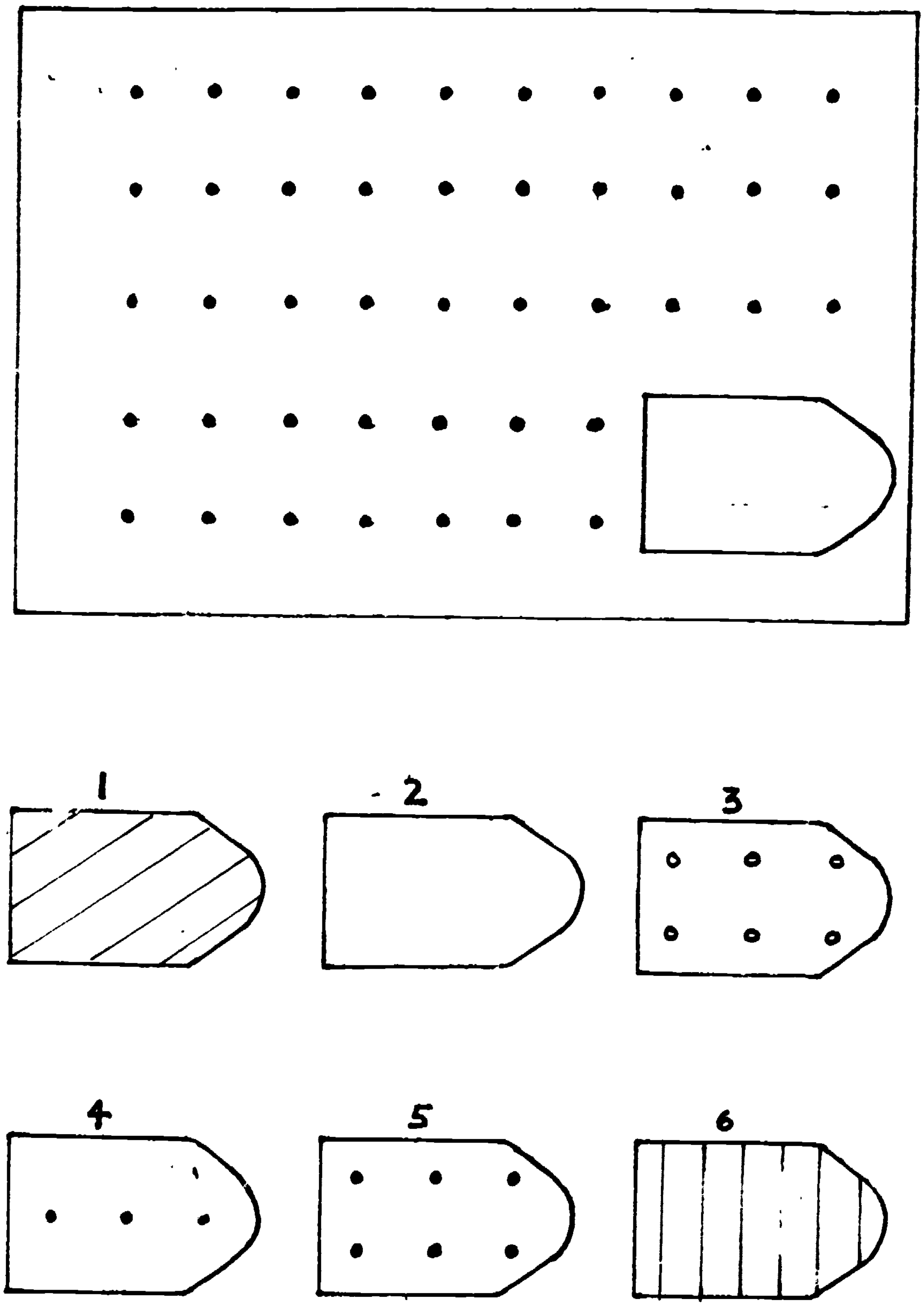
Let's look at No. 7 now. In this one we can again separate foreground and background. The backgrounds stay the same across the rows so we will look for a solution which has a kind of graph-paper background. The foreground, in No. 7, is either a square, an X or a dot and there's one of each in each row, and in each column (check by pointing). So what is missing in the last row: a square, an X, or a dot? An X, right. So the answer was No. 6.

Now No. 10. Do you see that by pushing together the first two figures in the top tow (pointing) we would get the third figure? It would also work for the second row, wouldn't it? So the answer is 8 because that is what we would get if we pushed the first 2 figures of the last row together. We should check this answer to see that it works for the columns too. It does.

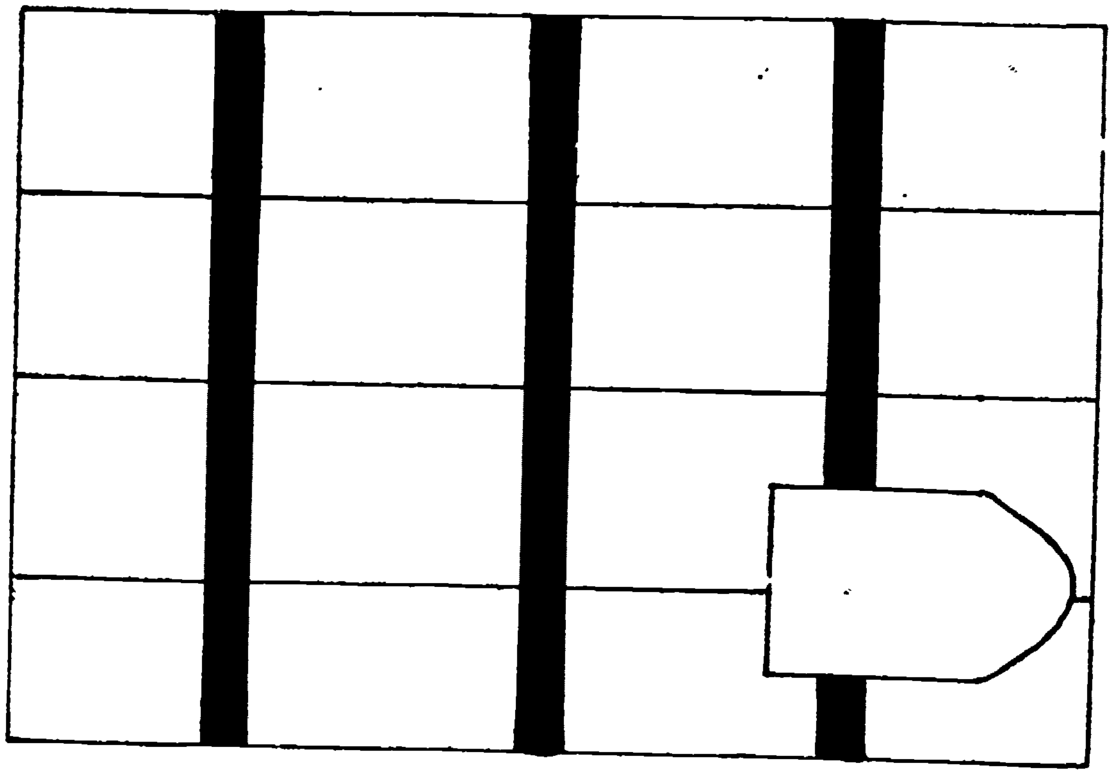
In No. 11, we can again think of the foreground and background separately. The foreground adds up along the rows, two dots and two dots make four dots (pointing) and the background adds up down the columns. The correct answer is 7. Always check both the *rows* and the *columns*. No. 12 was the most difficult. It can be thought of as a subtraction. Look at the middle column: if we take the second figure, the dot, away from the first figure, we'd be left with the third figure. This works for the first column too and also for the rows. So the correct answer is 6. The problem could also have been looked at as addition - add the second and third to get the first. There are often several ways to look at a problem, but always check that your answer works for rows *and* columns.

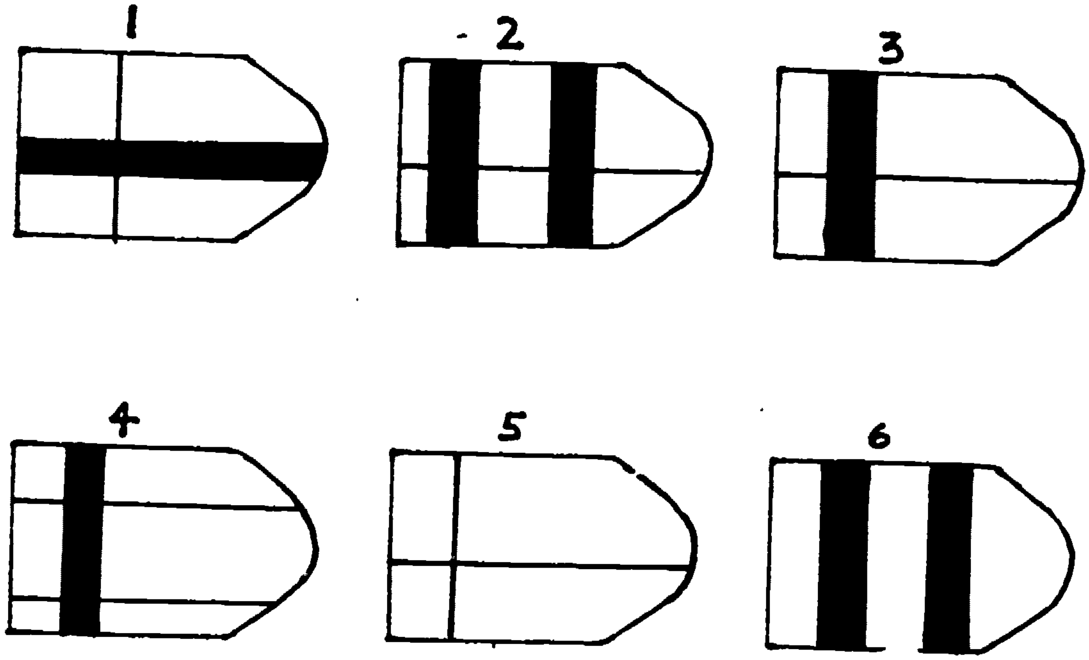
The second booklet, set II, has the same kind of questions. There are 36 of them. You can take as long as you wish. Please mark down the time you start and finish.

X1

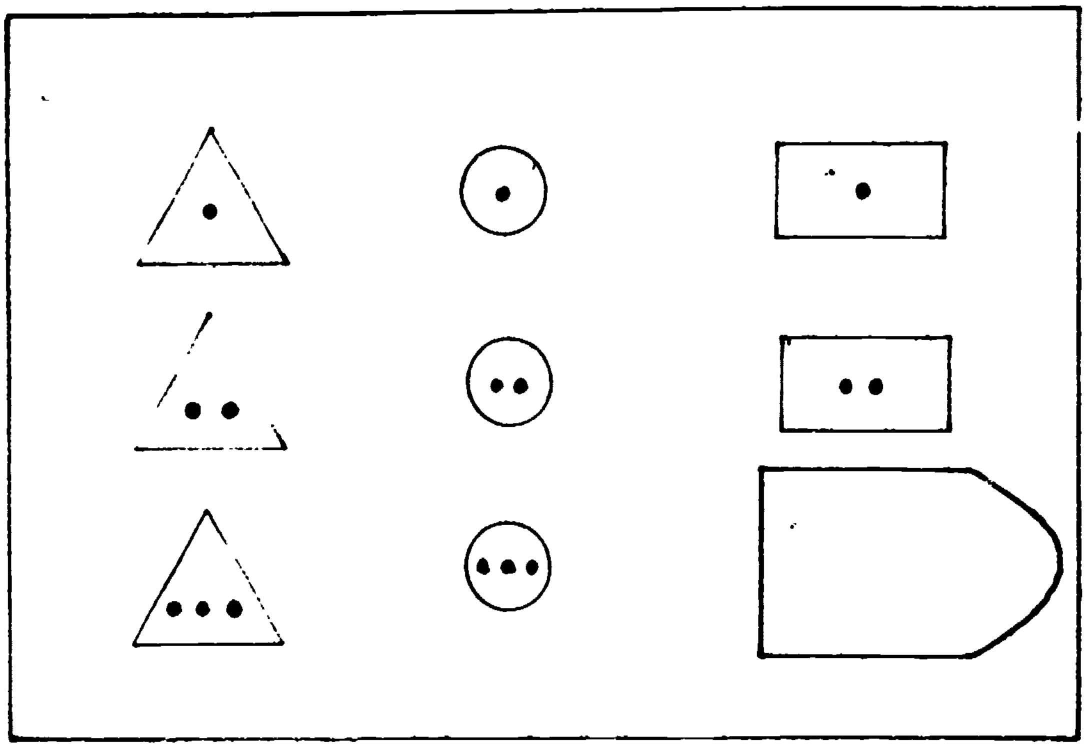


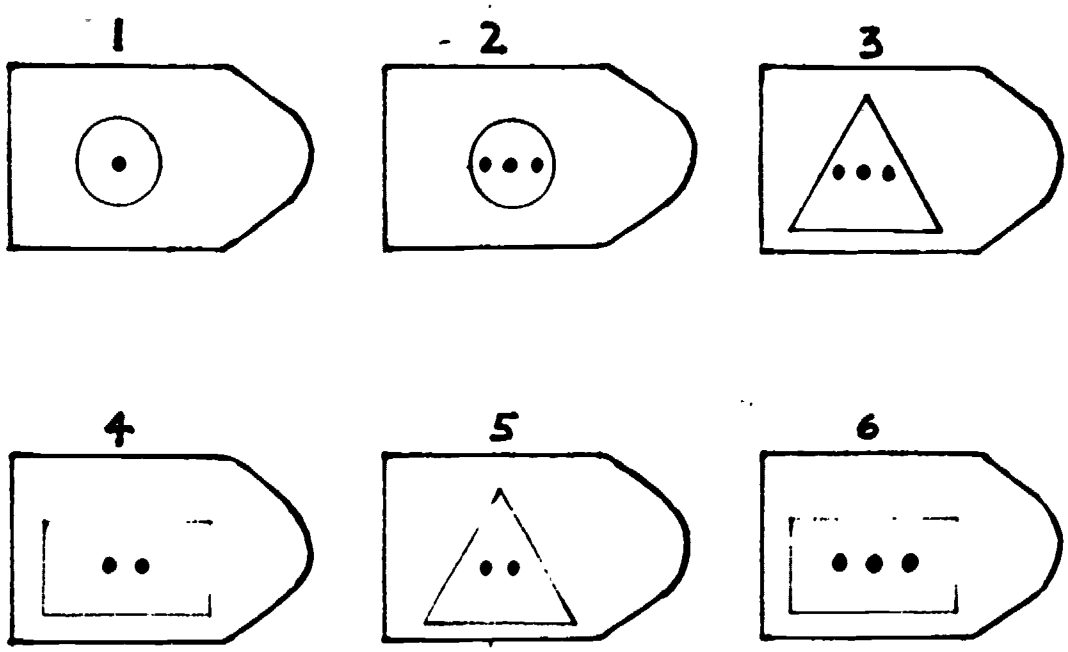
X2



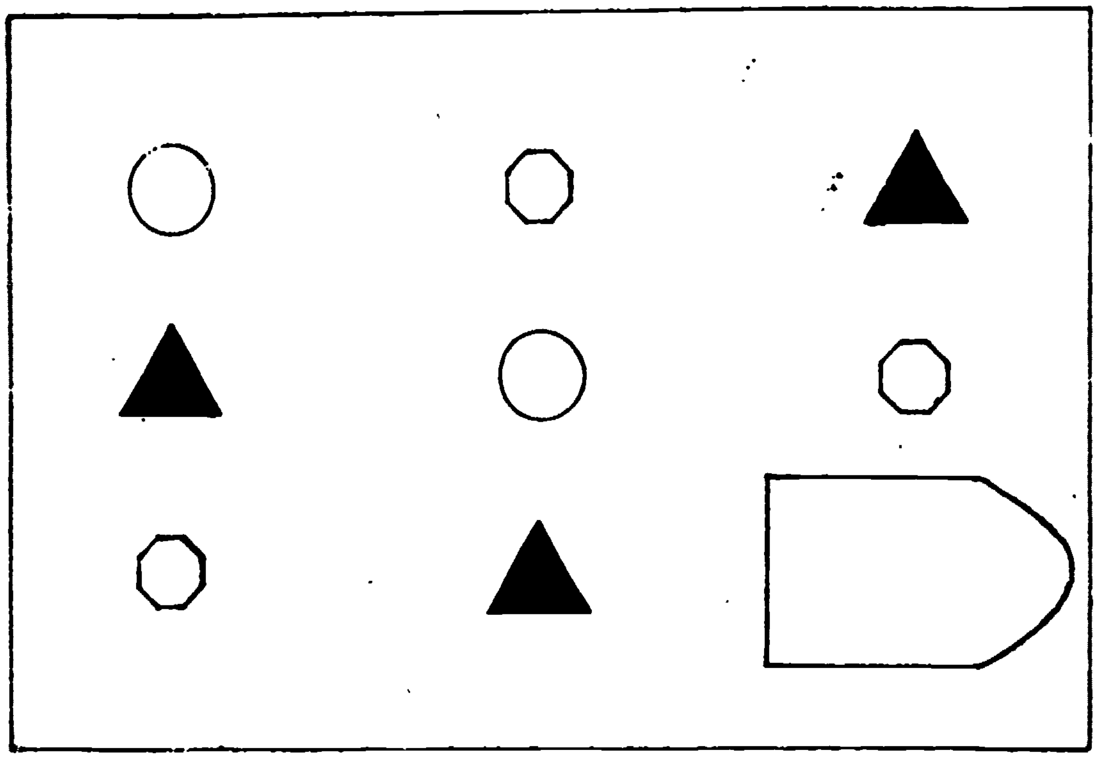


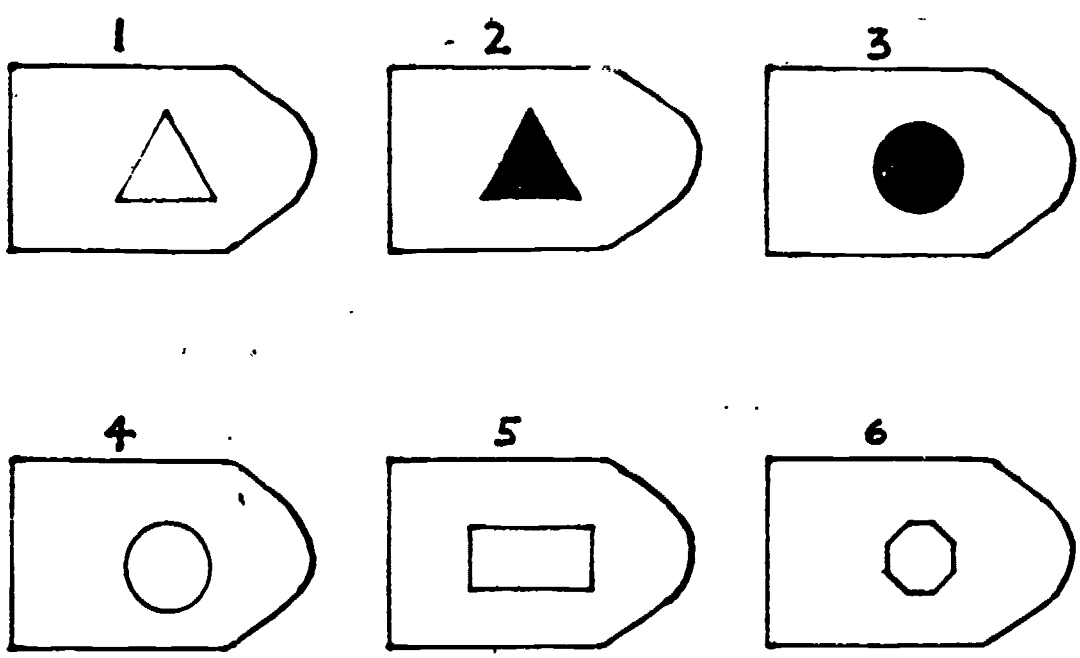
X3





X4





APPENDIX D

AN ATTEMPTED CROSS-VALIDATION

Strong relationships can occur by chance, particularly when the number of observations is small. It is therefore essential to apply the measures to a different sample to see if the strong relationship is maintained, i.e., to cross-validate.

Although funds remained for only eight more WISCs, it was decided to apply the recommended procedure to the eighth grade of a different school.

Sample used in cross-validation. Two schools, school Y and school Z, close to the original school were available. School Y is on the borders of a white community and was one of the last schools in the district to become a predominantly minority school. The other school, Z, is located on the borders of an inner-city depressed area, and all the elementary schools which feed into it receive title I funds. School Z was chosen for the cross-validation study. It was comparable to the original school in ethnic composition of the student body but smaller and located in a more depressed area.

Procedures for cross-validation. The administration of the SPM was described to the two eighth grade science teachers in a brief after-school meeting. They administered the tests to each of their eighth grade classes.

Papers from the 209 students were screened and about 30 students who scored over 45 were located. Sixteen scored above 48 and were given the APM under the same conditions as in the original study.

Papers had been screened down to a score of 45 in case the scores in this school were generally lower. It was encouraging that the high scores compared well with the scores from the original school.

Out of the top ten students (all of whom scored 21 or higher on the APM and 51 or higher on the SPM) eight were scheduled for testing with the WISC, but one earned the distinction of being the only student in the entire study who failed to appear to take the WISC. (All WISCs were given on a Saturday at the student's school.)

Results. Table 24 shows the scores on the SPM, APM and, where available, the WISC. Whilst the matrices scores were as high as in the original study, the WISC scores were lower. It had been hoped that half of those tested would score at or above 114 on the performance scale and so be comparable to the WISC performance criterion group identified at the original school. Four of the seven tested scored 110 or better and if the standard error of estimate of the WISC is about five points (Anastasi 1961) there is a twenty percent chance that a person with a true score of 114 receives a score of 110 or less. However, the fact remains that all the performance scale scores were below 114. How is this fact to be interpreted?

TABLE 24

SPM, APM AND WISC SCORES FOR THE CROSS-VALIDATION SAMPLE

			WISC		
Student	SPM	APM	Performance	Verbal	Full-Scale
ZA	55	28	110	104	107
ZB	54	20	NT		
ZC	52	25	107	96	101
ZD	52	23	111	104	108
ZE	52	21	108	104	107
ZF	51	23	110	92	101
ZG	51	24	NT		
ZH	50	23	NT		
ZI	50	23	110	97	104
ZJ	50	21	99	97	98
ZK	50	17	NT		
ZL	49	19	NT		
ZM	49	18	NT		
ZN	49	15	NT		
ZO	48	19	NT		
ZP	48	18	NT		

NT = Not tested

Possibly the screening and selection procedure is poor and students who would have scored over 1 14 were not selected for testing. Further testing at the same school would be needed to verify this and the implication would be that the results at the first school were due to very odd coincidences.

On the other hand several other explanations can be offered for the lower WISC performance scores despite the equally high matrices scores.

Firstly, there was a change in psychologist and, therefore, a possible examiner effect. The school psychologist who gave the original tests, a white male, is an experienced counselor of teenagers, reputed to communicate very effectively. Unfortunately he was not available for the cross-validation study. The school psychologist who gave the cross-validation tests, a white female is not yet so experienced. As an indication that the students were not entirely at ease, she reported that only one student accepted any of the candy she made available.

Another factor might have been the lower SES level of school Z. It might be hypothesized that the WISC, which involves the child's dealing with time pressure and a strange adult, is more affected by SES level than the matrices.