DOCUMENT RESUME

ED 060 595 EC 041 545

AUTHOR Findley, Warren G.; Bryan, Miriam M.
TITLE Ability Grouping: 1970 - Status, Impact, and Alternatives.
INSTITUTION Georgia Univ., Athens. Center for Educational Improvement.
SPONS AGENCY Bureau of Elementary and Secondary Education (DHEW/OE), Washington, D.C.
PUB DATE 71
NOTE 98p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Ability Grouping; *Grouping (Instructional Purposes); Minority Groups; Research Reviews (Publications); *Special Classes; *Student Placement; Surveys; Testing; *Test Validity

ABSTRACT

The examination of the use of ability grouping of students begins with presentation of the questionnaire responses from 328 school districts concerning how and how much ability grouping is practiced within their systems, on what basis students are assigned to groups, and how many poor or non-white students are involved. Following is a summary of research relevant to the impact of ability grouping on school achievement, affective development, ethnic separation, and socioeconomic separation. Consideration of the problems and utilities involved in the use of tests for grouping children with limited backgrounds focuses on test reliability and validity, cultural bias, publishers' test information, and use of tests with disadvantaged and Mexican American groups. The final section contains a series of brief accounts of alternative strategies to ability grouping. (KW)

The work presented herein was performed pursuant to a grant from the U.S. Office of Education, Bureau of Elementary and Secondary Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement by the U.S. Office of Education should be inferred.

Additional copies of this document are available upon request. Write:

Dr. Morrill M. Hall, Director
Center for Educational Improvement
College of Education
University of Georgia
Athens, Georgia 30601

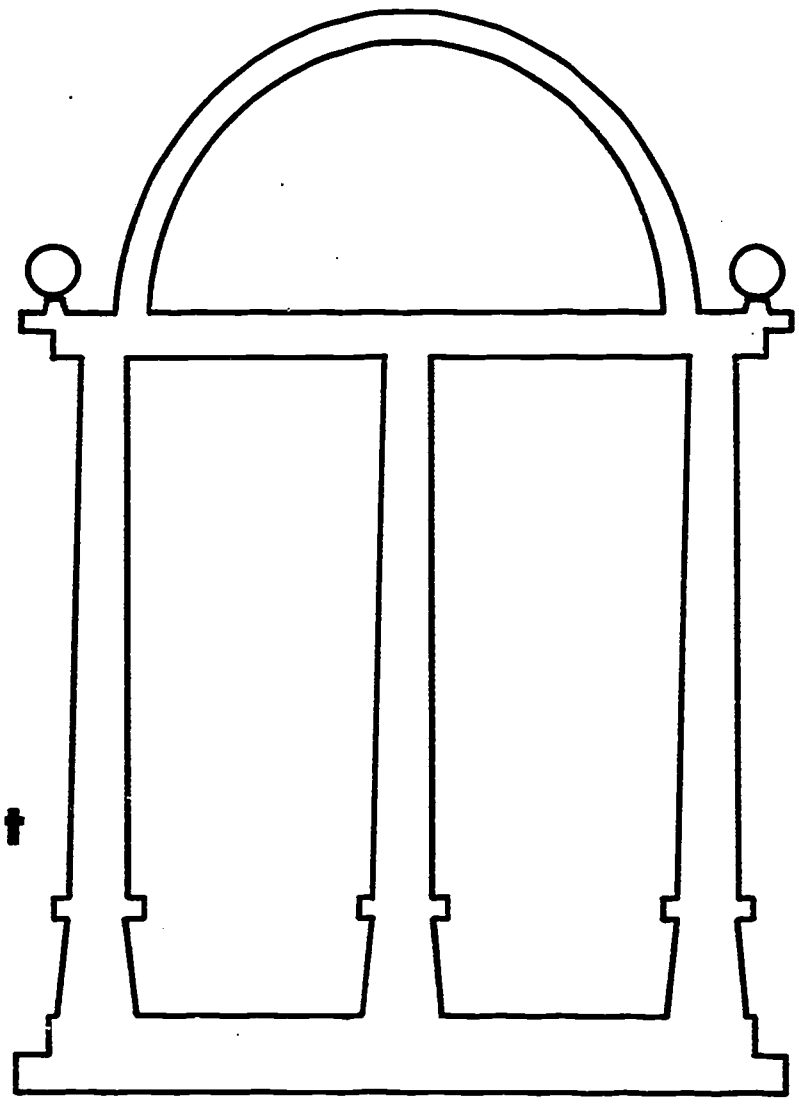
CEI

Ability Grouping: 1970

Status, Impact, and Alternatives

WARREN G. FINDLEY
and
MIRIAM M. BRYAN

Center for Educational Improvement
University of Georgia
Athens, Georgia 30601



FOREWORD

In December, 1969, a task force was organized for the purpose of advising on the scope and organization of a series of reports regarding ability grouping in the public schools of the United States. Those involved in the planning included:

Warren G. Findley, Principal Investigator

Miriam M. Bryan	Edmund W. Gordon
Paul I. Clifford	Roger T. Lennon
John E. Dobbin	A. John Stauffer
Gordon Foster	Ralph W. Tyler

The Office of Education and the U.S. Department of Health, Education, and Welfare were represented by Peter Briggs, Christopher Hagen, and Rosa D. Wiener.

Four documents were planned and, now completed, constitute the four sections of this report:

I. Common Practices in the Use of Tests for Grouping Students in Public Schools.

II. The Impact of Ability Grouping on School Achievement, Affective Development, Ethnic Separation, and Socioeconomic Separation.

III. Problems and Utilities Involved in the Use of Tests for Grouping Children with Limited Backgrounds.

IV. Alternative Strategies to Ability Grouping.

Mrs. Bryan prepared Section I, based on questionnaire responses from schoolmen and supplementary data from Miss Wiener. Dr. Clifford and Dr. Dominick Esposito prepared the basic content of Section II, which was then edited by Mrs. Bryan. Contributions to Sections III and IV were secured from Mrs. Bryan, Mr. Dobbin, Dr. Findley, Mrs. Blythe Mitchell, and Dr. Stauffer. The introductory section, giving a brief summary and highlighting the conclusions and recommendations, was prepared by Dr. Findley. As work progressed, Mrs. Bryan took fundamental responsibility for preparing tentative final drafts for the first three sections, verifying all information reported. She also participated with the Principal Investigator in decisions regarding the final drafts of all parts.

Special thanks go to the individual members of the task force for comment and criticism, especially in the early stages. Finally, very special thanks go to Dr. Morrill M. Hall, Director of the Center for Educational Improvement in the College of Education at the University of Georgia, for his unfailing support of this project at every stage.

January 1971 W. G. F.

TABLE OF CONTENTS

HIGHLIGHTS—CONCLUSIONS AND RECOMMENDATIONS

This is a summary in non-technical language of related information in the supporting sections. It summarizes them in a sequential series of statements that follow. If these are read in sequence, they form a logical argument or brief in support of the recommendations.

A few preliminary statements will help make the meaning of the conclusions clearer. Conclusions are to be read in the light of the general notion that effects are more favorable or less damaging as one progresses from situation D1 to situation D4 defined below.

Preliminary Statements

A. As used here, ability grouping is the practice of organizing *classroom groups* in a graded school to put together children of a given age and grade who have most nearly the same standing on measures or judgments of learning achievement or capability.

B. Grouping and regrouping *within a classroom* for instruction in particular subjects is an accepted and commended instructional practice. It is *not* to be considered ability grouping in the sense in which that term is used here.

C. Ability grouping may be based on a single test, on teacher judgment, or on a composite of several tests and/or judgments.

D. Ability grouping in a school district may take one of several forms, but chiefly one of four varieties:

1. Ability grouping of children in all school activities on the same basis.

2. Ability grouping for all learning of basic skills and knowledge on the same basis, but association with the generality of children of the same age in physical education and recreation.

3. Ability grouping for learning of basic *academic* skills and knowledge on the same basis, but association with the generality of children of the same grade in less academic activities, including physical education, art, music, and dramatics.

4. Ability grouping for learning of individual subjects or related subjects on different bases related to progress in mastering the different areas (for example, language arts vs. mathematics), but association with the generality of children of the same grade in non-academic areas. This has sometimes been referred to as "achievement grouping."

E. Ability grouping in the first grades, usually the first six or eight grades, is generally by assignment to single classroom teachers for instruction in most subjects.

F. Ability grouping in the last grades, usually in junior and senior high school, is generally by assignment *within* programs of study (college preparatory, commercial, vocational, general).

G. At high school, assignment to a curriculum or program of study may be made a part of a total ability grouping program. On the other hand, ability grouping is often accomplished to a degree by a process of self-selection in which individual students choose their programs of study freely or with some regard to prerequisites. In essential respects, the difference between the two methods is analogous to the distinction between *de jure* and *de facto* segregation.

H. Ability grouping practices differ in the degree to which reclassification or reassignment is provided for. Practices vary from virtually no review to systematic review at specified intervals of years or more often.

I. Ability grouping may be limited to provision for extreme groups.

J. Special education for mentally retarded children is to be distinguished from general ability grouping, but needs to be considered a special case subject to examination and report here.

K. Provision of advanced subjects for limited numbers of superior students is to be distinguished from ability grouping applied to all students of a grade group, but needs to be considered a special case subject to examination and report here.

Conclusions

1. Ability grouping is widely practiced in American school systems.

2. Ability grouping is especially characteristic of larger school systems.

3. Ability grouping is more common in higher grades than in earlier grades.

4. Homogeneous grouping by ability across the subjects of the school curriculum is impossible. Groups homogeneous in one field or sub-field will prove heterogeneous in other fields. Thus, children grouped by reading score or "intelligence" will overlap considerably in mathematics achievement.

5. Ability grouping is widely approved by school teachers and administrators.

6. Although unqualified approval of ability grouping is widespread among teachers, disproportionate numbers express preference for teaching mixed, average, or superior classroom groups over teaching lower-achieving groups.

7. Substantial educational research on streaming

(homogeneous grouping) in England's schools indicates that the most detrimental effect is caused by assigning "prostreaming" teachers to "non-streamed" classes. The generalization also applies to American schools.

8. Socioeconomic and social class differences are increased by streaming, reduced by non-streaming.

9. Virtually all ability grouping plans depend on tests of aptitude or achievement as an integral feature.

10. Ability grouping, as practiced, produces conflicting evidence of usefulness in promoting improved scholastic achievement in superior groups, and almost uniformly unfavorable evidence for promoting scholastic achievement in average or low-achieving groups. Put another way, some studies offer positive evidence of effectiveness of ability grouping in promoting scholastic achievement in high-achieving groups; studies seldom show improved achievement in average or low-achieving groups.

11. The effect of ability grouping on the affective development of children is to reinforce (inflate?) favorable self-concepts of those assigned to high achievement groups, but also to reinforce unfavorable self-concepts in those assigned to low achievement groups.

12. Low self-concept operates against motivation for scholastic achievement in all individuals, but especially among those from lower socioeconomic backgrounds and minority groups.

13. Children from unfavorable socioeconomic backgrounds tend to score lower on tests and to be judged less accomplished by teachers than children from middle-class homes. This discrepancy is more marked as children grow older and approach adulthood.

14. The effect of grouping procedures is generally to put low achievers of all sorts together and deprive them of the stimulation of middle-class children as learning models and helpers.

15. Low achievers include many disruptive children who have failed to acquire constructive school attitudes as well as children with low and slow achievement patterns.

16. Children of many minority groups (Negro, Puerto Rican, Mexican-American, Indian American) come disproportionately from lower socioeconomic backgrounds.

17. The source of disadvantage for some minority groups (Puerto Rican, Mexican-American, Indian American) derives in part from the fact that teaching and testing in schools are usually entirely in English, which for them is a "second" language.

18. The language patterns of black and white children from lower socioeconomic backgrounds often differ so markedly from "standard American" as to make schooling in most schools involve language disability by such language standards. This circumstance has not only the direct effect of making learning more difficult. Indirect effects are also produced via lowered self-concept because of frequent corrections.

19. A fundamental generalization is that differences in socioeconomic backgrounds result in cumulative effects because of early acquired differences in ability to interact profitably with teachers who have middle-class habits and values. Middle-class children come to school prepared to respond to approval by teachers for their prior learning and readiness to respond. Disadvantaged children, especially boys, often have to unlearn assertive, unresponsive behavior in order to participate in a teaching-learning rapport in the classroom.

20. Desegregated classes have greatest positive impact on school learning of socioeconomically disadvantaged children when the proportion of middle-class children in the group is highest. Conversely, when socioeconomically disadvantaged children are in the majority in a class, the effect of grouping is commonly to produce poorer achievement on their part.

21. Assignment to low achievement groups carries a stigma that is generally more debilitating than relatively poor achievement in heterogeneous groups.

22. A positive dynamic of all instructional programs is constructive stimulation, what J. McV. Hunt calls "the problem of the match"—some stimulation, but not too much, accompanied by supportive encouragement.

23. Formal education, or instruction, makes a difference in ultimate adult capability. How much difference education makes in comparison with other factors is a separate question which is essentially irrelevant.

24. Ability grouping practices are to be distinguished from each other in terms of their underlying strategies for dealing with initial differences among children and the cumulative effect of such differences.

25. Different ability grouping practices show different amounts of differential treatment given to different children after ability grouping has been done. The teaching strategies employed with those classified low often deny stimulation offered to those classified high on the criterion used in grouping. Elsewhere, all those classified in one group are thereafter taught as if almost identical in capability.

26. Of the patterns of ability grouping differentiated in Preliminary Statement D, type D4 generally involves more detailed diagnosis and specific instructional differentiation.

27. There are viable alternatives to ability grouping as means of furthering school learning, including

stratified heterogeneous grouping, tutoring, team teaching, and individually programed instruction.

28. Planned heterogeneous grouping—notably the Baltimore plan of stratified heterogeneous grouping by tens—takes into account simultaneously the concern for curtailing extreme heterogeneity, while assuring enough diversity to give leadership opportunities in each class, providing thereby for stimulation of the less advanced by these leaders, and avoiding the concentration of defeated and stigmatized children in a bottom group almost impossible to inspire or teach.

29. Where older children, themselves academically retarded, are paid to tutor younger children who are having difficulty in learning to read in the elementary grades, both groups gain substantially. In fact, the older children gain even more than the younger ones being tutored. Similar findings apply to writing.

30. Teaching by teams of teachers with different responsibilities, under the leadership of coordinating master teachers, is a fundamental pattern in plans developed for training future elementary school teachers. Departmentalization of instruction may be considered a step in this direction.

31. Individualized instruction by prescription of sequences of learning experiences has been worked out for much of the learning of basic skills and structured knowledge.

32. All four of the above teaching-learning practices can be applied simultaneously. They are mutually compatible.

33. Early childhood education, whether designed to be compensatory or for all children, presents a further supplementary approach.

34. Residential segregation, in the form of concentrations of minority groups in cities and the moving of majority groups to suburbs, plus the organization of private schools along ethnic lines, makes ethnic desegregation within many large cities almost meaningless.

35. The same may be said to a lesser degree of socioeconomic segregation without regard to ethnic distinctions.

36. Ability grouping of the types described in Preliminary Statements D1—D3 has generally undesirable effects on learning and self-concept within like ethnic and socioeconomic groups, which are magnified when the correlated factors of ethnicity and socioeconomic status are involved.

37. Findings of the impact of ability grouping on classroom groups have implications for residential segregation and schooling tied to it. The issues underlying ability grouping and school desegregation are deeply embedded in our society and its culture. The matters reported here are integral parts of a larger social pattern, contributing to the perpetuation or change of that pattern, but largely determined by it.

Recommendations

1. Ability grouping of the types described in Preliminary Statements D1, D2, and D3 should not be used.

2. Ability grouping of the types described in Preliminary Statement D4 may be used to advantage where the information gained by testing and/or observation is the first step in a program of diagnosis and individualized instruction.

3. Provision should be made for frequent review of each individual's grouping status as part of the instructional program.

4. Tutoring, team teaching, individually programed instruction, and early childhood education should be explored and exploited for their usefulness in promoting learning.

5. The personality dynamics of the tutoring of younger children by older children, often of modest ability, should be explored and exploited.

6. Heterogeneous grouping, in a classroom atmosphere of cooperation and helping, should be the rule except as indicated under Recommendation 2.

7. Stratified heterogeneous grouping by tens, as practiced in Baltimore, should be utilized and refined.

8. Favorable self-concept should be a goal in itself, but it is also a supportive factor in learning. An attitude of firm confidence and hope by the teacher is fundamental. Techniques for conveying such an attitude can be learned.

9. Teacher training should include an emphasis on welcoming diversity in children, and teaching children to prize it in each other. A particularly important aspect of such diversity is with regard to language and customs of minority groups. Teachers therefore need pre-service and/or in-service preparation in language habits and cultural heritages of minority groups to use as the basis for positive acceptance of all kinds of children into the classroom group.

10. Steps should be taken *as early as possible* in each local situation to promote unitary school populations in each district and each classroom. When a district or city has become almost completely a socioeconomically limited population, the possibility of effective desegregation and its constructive impact virtually disappears.

I. COMMON PRACTICES IN THE USE OF TESTS FOR GROUPING STUDENTS IN PUBLIC SCHOOLS

HISTORY AND PREVIOUS STUDIES

Grouping in both elementary and secondary schools has been a topic of perennial interest in the United States for about a hundred years. The origins of grouping actually go further back than that—to the middle of the nineteenth century, when growing numbers of children in school began to result in change, first, from the ungraded, one-room, one-teacher school to the primary-intermediate or two-room, two-teacher school and, finally, to graded, many-room, many-teacher schools with their consequent reduction in the range of differences in age and academic ability within each classroom.

The reduction of differences was, however, not great enough to prevent a high failure rate in single-grade classrooms, where emphasis now was being placed on the mastery of subject matter with steady progress from grade to grade. In the face of adverse reaction from both without and within the school to the retention of large numbers of older children in the elementary grades, educators began to look for ways of individualizing instruction so that school work could be completed at a different rate by each student.

A number of approaches to individualized instruction were developed and carried out between 1890 and 1910, and much research was built around them, but no conclusive evidence was ever obtained to show that they were particularly effective educationally. Teachers were overwhelmed by the problems that wide ranges of intellectual ability among students of the same age presented for a program of individualized instruction, and large numbers of students continued to fail the strictly subject-matter oriented courses of study.

Immediately following World War I, attention turned to the possibility of using group intelligence tests of the type developed during the war to measure learning ability and to form ability groups on the basis of test results. Scores on group intelligence tests and, a few years later, on standardized achievement tests became the measures on which were based most of the grouping practices between 1920 and 1935.

As a result of evidence offered by numerous research projects during this period, which failed to show that students grouped on the basis of scores on either intelligence or achievement tests were able to achieve greater subject-matter mastery than were students in heterogeneously grouped classrooms, and as a result, too, of the opposition of the proponents of progressive education to what they considered to be an undemocratic form of school organization that stigmatized slower students and made snobs out of the abler ones, ability grouping went into a period of relative decline.

From 1935 to 1950, the amount of ability grouping practiced was considerably less than that of the earlier 15-year period, and ability grouping was not a particularly popular topic for research. School people who continued to employ ability grouping because it was administratively convenient and popular with teachers, and with some parents and students, had to admit that, despite efforts to improve their grouping procedures, students grouped on the basis of IQ or level of achievement still presented a wide range of differences in ability to learn generally and in ability to perform uniformly well or at the same speed in all subjects.

During the past 15 years, since the middle 1950's, there has been renewed interest in ability grouping—and a number of different patterns have emerged. For one thing, there is somewhat more concern today than formerly with special education for the gifted, with some impetus here undoubtedly the result of the launching of Sputnik and the consequent emphasis on special training for students with talents in mathematics, science, and foreign languages; at the other end of the intellectual scale, children who present special problems of educability because of mental retardation, physical handicaps, or cultural deprivation have been given more special attention than previously. Some schools have gone still further and differentiated among high average, average, and low average students.

While relatively limited quantitative information has been available in recent years regarding grouping practices, at least three fairly thorough surveys have been reported:

The NEA Research Division in 1962 reported that during the school year 1958-59, 77.6 percent of 3,418 school districts 2,500 and over in population were making some use of ability grouping in the elementary grades, and that 90.5 percent of these districts were using it at the secondary school level. Of the districts reporting, 51.7 percent said they planned to add or expand ability grouping in the elementary grades, and 67.3 percent said they planned to add or expand it at the secondary school level. Fewer than one percent indicated plans to curtail ability grouping.

During the 1960-61 school year a study of grouping in early elementary education was conducted by the

U. S. Office of Education. Assignment of children to kindergarten classes on a homogeneous basis or on a partially homogeneous basis was reported by 6.6 percent and 14.7 percent, respectively, of the 5,559 districts responding, while 78.7 percent of the districts reported heterogeneous grouping at this level. By the third grade, 15.8 percent of 10,608 districts reported homogeneous grouping and 33.5 percent partial homogeneous grouping, with 50.7 percent of these districts still reporting a policy of heterogeneous grouping. Thus, the shift to homogeneous grouping was found to be well under way at the end of the primary level.

Data obtained from a questionnaire on administrative practices within the elementary school, distributed by the NEA Research Division to a sample of school systems in early 1966, showed 24.9 percent of the 12,130 schools reporting to be assigning children to classes on a random basis, 43.2 percent to be specially grouping a few children but not most, and 27.5 percent carefully grouping all children, while 4.4 percent gave no indication. The heaviest emphasis on the careful grouping of children was reported by school systems with enrollments of 100,000 or more (45.8 percent).

It should be noted that the recent trend in the direction of the increased use of ability grouping has taken place in the face of newer and steadily increasing evidence from research study after research study that the various patterns of ability grouping tend to show little or no significant increase in achievement for children at any intellectual level and no little damage to the other aspects of the development of the children involved.

THE QUESTIONNAIRE STUDY

In an effort to get as much up-to-date information about grouping practices as could be gathered, it was decided to solicit the help of state school officers, directors of research in large cities, and individuals known to be concerned with research studies involving children of minority or other disadvantaged groups. Letters were addressed to all 50 state school officers asking them to identify school systems within their states in which ability grouping has been or is being practiced and from which information concerning grouping procedures and the advantages and disadvantages of ability grouping to the system might be obtained. Approximately 400 such school systems were identified and each of these was asked to complete the brief questionnaire appended to this section and to supply other printed or written data describing how current grouping procedures have developed and how they work. Letters addressed to directors of research in 77 large cities, virtually all cities of over 200,000, asked that the same questionnaire be completed by them and that reports of any research undertaken in their cities in which ability grouping was involved be made available to the committee. Finally, letters were directed to 15 individuals in various parts of the country, known to have been involved in research having to do with school problems of children of Negro, Mexican-American, or American-Indian parents, or of white children in families of low socioeconomic status, who might have useful information for the committee.

Of the replies received from research directors in large cities, 10 were from the Northeast, 18 from the South, 13 from the Middle West, 6 from the Southwest, and 11 from the West—various regions being made up of the states assigned to these regions in the Coleman report of the Educational Opportunities Survey.* Of the replies received from school administrators, 79 concerned schools or school districts in the Northeast, 47 in the South, 59 in the Middle West, 23 in the Southwest, and 62 in the West. Replies, then, were received from 328 individuals in all.

It should be pointed out here that the data requested were for school districts, not for individual schools. Data were supplied for systems with school populations ranging from more than 1,000,000 to fewer than 100. Since virtually every large city and several county systems responded as units to the questionnaire, it seems safe to say that the number of schools represented is well beyond 5,000.

Many local school officers supplemented the completed questionnaire with letters, pamphlets, and books describing in much more detail than was possible on the questionnaires the philosophy and practices of their districts with regard to grouping. Substantial printed documents are listed as supplementary references in the bibliography for this section. Of the school officers replying, only five wrote that the pressure of other activities would prevent their taking time to assemble the information necessary for completing the questionnaire.

---

*Northeast—Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania, Delaware, Maryland, District of Columbia.

South—Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi, Tennessee, Kentucky, Louisiana, Arkansas.

Middle West—Ohio, Indiana, Illinois, Michigan, Wisconsin, Minnesota, Iowa, North Dakota, South Dakota, Nebraska, Kansas, Missouri.

Southwest—Arizona, New Mexico, Oklahoma, Texas.

West—Montana, Idaho, Wyoming, Colorado, Utah, Nevada, California, Oregon, Washington, Alaska, Hawaii.

The replies to the first seven questions on the questionnaire are summarized for the five regions and for the country as a whole. A second table for question 1, Table 1b, summarizes the incidence of ability grouping in terms of size of school district for the first 308 districts reporting. Second tables for questions 6 and 7, numbered 6b and 7b, report the numbers of children represented in the school district totals reported in Tables 6a and 7a, respectively. The replies to questions 8 and 9 are summarized for four different groups of school districts: those employing grouping generally on a district basis, those employing grouping at some grade levels or in some subject matter areas, those in which grouping procedures and practices vary from school to school, and those not employing grouping either as a matter of district policy or on an individual school basis.

In interpreting the results of the questionnaire, three questions that might be asked of any individual making a self-report should be kept in mind:

1. Did the individual understand the question asked?
2. Did the individual know his school or school district sufficiently well to respond correctly?
3. Did the individual want to respond correctly?

There are reasons to believe that these questions cannot in all cases be answered in the affirmative. Certain questions were obviously misunderstood by some individuals completing the questionnaire. The nature of the response in other cases indicated that some individuals did not know their schools or school districts well enough to be able to supply the information requested. And the failure of some individuals to respond to certain questions may be interpreted as omission by design. Insofar as these conditions are present, a systematic error in information reported may exist. Entries in the tables indicating "Information Incomplete" reflect the extent of this defect quite accurately.

Question 1

Are students at any grade level in your school district grouped homogeneously?

If the individual completing the questionnaire answered question 1 with an unqualified "Yes" and indicated in response to question 2 that grouping was done in more than one subject or in more than one grade, the response was tallied as "Generally." Grouping for a single subject or for a single grade was tallied as "Partially."

As can be seen from Table 1a below, better than 55 percent of the school districts from which replies were received do some grouping in more than one subject or grade on a district-wide basis and approximately 77 percent do grouping of some kind. The percentages are not significantly different from those reported by the NEA Research Division in their 1962 summary.

Table 1b reports the use of grouping in terms of the size of the student population for 308 school districts. While the incidence of the grouping is slightly erratic, the tendency is in the direction of greater use of grouping in districts with larger school populations. The unusually large incidence of grouping shown in school districts with populations of less than 1,000 is largely the reflection of the wide use of ability grouping in small school districts in the Midwest, while the low incidence of grouping in the South and West influences the figures across the table.

A minor trend is for school districts with populations under 25,000 to do more "partial" grouping within schools, while those over 25,000 more frequently allow variation from school to school.

The single subject for which grouping was reported most frequently was reading, with mathematics in second place. With or without ability grouping by

Table 1a
Extent of Homogeneous Grouping, by Geographical Location

	Northeast	South	Middle West	Southwest	West	Total
Generally	61	26	40	14	39	180
Partially	10	11	10	1	3	35
Varies with School	5	9	11	5	7	37
Generally No, Unclassifiable	0	1	0	2	2	5
No Grouping	12	18	9	7	20	66
Not Able to Respond	1	0	2	0	2	5
Total	89	65	72	29	73	328

Table 1b
Extent of Homogeneous Grouping, by Size of School District

	Less than 1,000	1,000-5,000	5,000-10,000	10,000-25,000	25,000-50,000	50,000-100,000	100,000-500,000	More than 500,000	Total
Generally	15	41	33	39	16	17	8	2	171
Partially	6	10	7	9	1	1	1	0	35
Varies with School	0	3	2	4	8	9	6	1	33
Generally No, Unclassifiable	0	1	0	1	0	2	0	0	4
No Grouping	2	14	7	16	8	11	2	0	60
Not Able to Respond	0	1	0	2	0	2	0	0	5
	23	70	49	71	33	42	17	3	308

class, a large number of respondents reported that grouping for reading and mathematics was done within classes.

Several respondents reporting vertical grouping, either within grade or within class, emphasized that the grouping was flexible—that students could move from level to level upon meeting the criteria for a particular level. Others pointed out that grouping, especially at the elementary school level, was done by basic skill areas and that a student might be assigned to groups at different levels in different skills. Still others called attention to the fact that, unless students are locked into a tracking system. grouping at the secondary school level may be largely a matter of self-selection.

A considerable number of respondents indicated that homogeneously grouped classes had at some time recently been replaced by heterogeneously grouped classes, or were about to be, and that emphasis was being placed upon individualized instruction. Continuous progress concepts, computer-assisted instruction, team teaching, enrichment programs, and compensatory programs were mentioned as being employed with heterogeneous groups in the interest of better meeting the needs of the individual student.

Only two of the respondents now using heterogeneous grouping reported that their school districts were moving toward homogeneous grouping. One of these wrote:

> In the future we may have to consider grouping, especially in reading. As we move into the advanced stages of desegregation, it may be necessary to consider additional areas.

Question 2

If so, at what grade levels is homogeneous grouping done?

That practices regarding the grade levels at which homogeneous grouping is done vary widely is evident in the table on page 9, which shows the responses to question 2. As a matter of fact, even more variations were reported than are shown here, where only the grade levels at which homogeneous grouping is *mainly* done in any school district are indicated. Respondents reported different practices from school to school within district, different practices from grade to grade within school or district, and different practices for elementary, junior high, and senior high schools.

Of the 252 school districts reporting the use of such grouping on a systemwide basis, approximately 4 percent indicated that this was begun at the kindergarten level, while another 23 percent indicated that it was begun in Grade 1. (The response "All" has been interpreted here as grades 1 through 12 rather than grades K through 12.) In the 252 schools, approximately 29 percent of the students had been grouped by the end of Grade 3, 37 percent in two grades or more by the end of Grade 6, and 73 percent in one or more grades by the end of Grade 9. One hundred thirty-three, or 53 percent, of the respondents reporting the use of ability grouping indicated that the grouping, whether begun in primary, intermediate, junior high, or senior high school grades, continued through Grade 12.

No one of the respondents reported assignment to different schools on the basis of grouping. All were concerned with grouping within school, within subject matter area, or within class.

Table 2
Grade Levels at which Homogeneous Grouping Is Done

	Northeast	South	Middle West	Southwest	West	Total
All	10	4	6	1	6	27
K-12	2	1	2	1	5	11
1-12	6	4	2	0	0	12
1-3	0	0	2	1	2	5
1-4	0	0	0	0	2	2
1-6	2	0	1	0	0	3
1-8	3	3	1	1	0	8
1-10	0	0	1	0	1	2
3-12	2	0	0	0	0	2
4-6	0	1	2	0	2	5
4-9	0	1	1	0	1	3
4-12	1	4	1	0	0	6
5-8	1	1	0	0	0	2
5-9	2	0	0	0	0	2
5-12	1	0	1	0	0	2
7-8	6	0	3	0	0	9
7-9	6	0	2	1	4	13
7-12	15	8	5	5	8	41
8-12	2	0	8	0	1	11
9-12	4	3	3	4	4	18
10-12	2	0	1	0	0	3
Varies with School	5	9	11	5	7	37
Other	6	5	6	0	6	23
Information Incomplete*	0	3	2	3	2	10
No Grouping	12	18	9	7	20	66
Not Able to Respond	1	0	2	0	2	5
Total	89	65	72	29	73	328

*Includes 5 whose response to question 1 was recorded "Generally No, Unclassifiable."

Question 3

How long has homogeneous grouping been practiced in your district?

The information given in response to question 3, summarized in the table on page 10, is interesting because it reflects the uneven history of grouping. Fifty-one respondents, or 20 percent, indicated that homogeneous grouping had been practiced in their districts for 30 years or more, placing the introduction at some time during the years of early popularity of this kind of school organization. One respondent reported that homogeneous grouping had been practiced in his district since 1890, when such grouping was little more than an idea. Thirty four respondents, or 13 percent, reported the introduction of homogeneous grouping between 1940 and 1954, a period when grouping was at the nadir of its popularity. But 143 respondents, or 57 percent, reported its introduction during the past 15 years, when it has enjoyed a period of increasing

support by administrators and teachers in spite of the lack of conclusive evidence regarding its effectiveness in the improvement of learning.

Several respondents reported that grouping had been practiced in their districts for many years but in varying and continually changing ways to conform with new developments in educational theory and practice. Some indicated that the introduction of the ungraded primary school in recent years had been responsible for their currently grouping in the early grades; others reported that grouping had been recently introduced with the development of special programs for the academically talented and the mentally retarded.

Table 3
How Long Homogeneous Grouping Has Been Practiced in the District

Number of Years	Northeast	South	Middle West	Southwest	West	Total
1-5	10	8	14	3	7	42
6-10	13	12	19	3	14	61
11-15	13	8	5	6	8	40
16-20	10	2	2	3	4	21
21-30	4	2	5	2	0	13
30+	3	3	3	3	3	15
Many	11	4	7	0	6	28
Always	3	0	0	0	4	7
Varies With School	0	1	0	0	0	1
Information Incomplete*	9	7	6	2	5	29
No Grouping	12	18	9	7	20	66
Not Able to Respond	1	0	2	0	2	5
	89	65	72	29	73	328

*Includes 5 whose response to question 1 was recorded "Generally No, Unclassifiable."

Question 4

On what basis are your students assigned to homogeneous grouping? (If on the basis of test scores, please name the test.)

The information provided in response to question 4 leaves little doubt that test scores play a major role in group assignments, whether by themselves or in combination with other criteria. As is shown in the table on page 11, 206 of the 252 school districts reporting the use of homogeneous grouping, or approximately 82 percent of these districts, use test scores as the basis, or as one of the bases, for group assignments.

The information provided in the table must be interpreted with considerable caution since the question did not require school districts to report how highly structured were the procedures for assigning students to groups or, when multiple criteria were given as the basis for making group assignments, how the different criteria were weighted. Some respondents did, however, provide detailed information about their grouping procedures and others indicated the order of importance given the different criteria in reaching decisions regarding group assignments.

An examination of the information provided indicates that in some school districts grouping is done according to a highly structured, district-wide plan that varies only from elementary to junior high to senior high school. In other districts the procedures vary from school to school with the local faculties responsible for determining them. Several districts with highly structured procedures for grouping describe these in detail in printed booklets available to teachers, parents, and other interested persons.

If one can assume that multiple criteria listed by the respondents were given in the order of the relative

Table 4
Basis for Assigning Students to Homogeneous Groups

	Northeast	South	Middle West	Southwest	West	Total
Test Scores Only	7	7	9	2	8	33
Test Scores and School Grades	9	3	4	3	3	22
Test Scores and Teacher, Counselor, and/or Principal Judgment	18	13	16	5	17	69
Test Scores, School Grades, and Teacher Judgment	8	3	5	2	1	19
School Grades, Teacher Judgment, and Student Interest	1	1	7	2	3	14
Many Criteria (Test Scores, Teacher Judgment, Grade Averages) Plus Student and/or Parent Desire	23	12	16	5	5	61
Miscellaneous Single Criteria	10	4	3	1	8	26
No Specific Criteria—Varies with Local Practice	1	1	1	0	3	6
Information Incomplete*	0	2	0	2	3	7
No Grouping	12	18	9	7	20	66
Not Able to Respond	0	1	2	0	2	5
Total	89	65	72	29	73	328

*Includes 5 whose response to question 1 was recorded "Generally No, Unclassifiable."

weights assigned them, then test scores, school grades, and teacher judgment are generally considered to be the most important criteria, with approximately equal numbers of districts placing each of these at the top of the lists provided. Most respondents who did indicate an order of importance for different criteria reported that group assignments were made chiefly on the basis of teacher judgment and past performance, with test scores used principally to substantiate teacher judgment. A single, large city in the Northeast reported that group assignments were the responsibility of the school principal, the only directive from the central office being "that students are not to be grouped on the basis of a single test score alone."

More than 50 different standardized tests were identified by the respondents as being used in their districts. Ranking highest among these in terms of use are the following:

Readiness—*Metropolitan Readiness Tests*

Achievement—*California Achievement Tests, Iowa Tests of Basic Skills, Iowa Tests of Educational Development, Metropolitan Achievement Tests, Stanford Achievement Test*

Aptitude—*Differential Aptitude Tests*

Intelligence—*Lorge-Thorndike Intelligence Tests, Otis-Lennon Mental Ability Test*

These and some of the other widely used tests are given special attention in the third section, in which the problems and utilities of tests used for grouping are treated.

Question 5

How many students in all are involved in your homogeneous grouping plan?

As indicated in the table below, useful information was obtained from 207 of the school districts in which homogeneous grouping is practiced. More than 30 respondents reporting district-wide grouping or the percent of students involved in grouping did not give school enrollment figures for the district; 28 respondents replied that the number of students involved in their grouping plan was not known; and nine respondents chose not to answer the question at all. The assistance of the U.S. Office of Education was solicited in obtaining total enrollment figures for all districts involved. Combining this information with the figures supplied by respondents made it possible to reduce the number of responses that could not be used to 45.

It is interesting to note that while 67 districts with school populations of 25,000 or over reported that homogeneous grouping, generally or partially, was practiced in their districts as a matter of district policy (see Table 1b), only 20 of these districts reported the involvement of 25,000 or more students in their grouping plan. This is to a large extent the result of grouping at selected grade levels rather than at all grade levels. That practices vary widely in this regard was noted earlier.

Table 5
Numbers of Students Involved in Homogeneous Grouping

	Northeast	South	Middle West	Southwest	West	Total
Less than 2,500*	31	10	24	7	25	97
2,500-5,000	15	7	9	5	8	44
5,000-10,000	9	4	5	2	2	22
10,000-25,000	6	9	3	2	4	24
25,000-75,000	1	2	3	1	0	7
75,000-125,000	2	2	0	0	0	4
125,000-200,000	4	1	0	0	0	5
More than 200,000**	1	0	1	0	2	4
Information Incomplete***	7	12	16	5	10	50
No Grouping	12	18	9	7	20	66
Not Able to Respond	1	0	2	0	2	5
Total Number of Districts	89	65	72	29	73	328
Total Number of Students Involved	1,850,240 +	541,272 +	575,883 +	102,105 +	793,634 +	3,863,134 +

*Several school districts reported grouping in a single subject or at a single grade level.

**Two large city school systems reported grouping for 750,000 and 553,338 students, respectively.

***Student populations of these school districts were known, but not the number of students involved in homogeneous grouping. Includes 5 whose response to question 1 was recorded "Generally No, Unclassifiable."

Question 6

What percent of these students are from low socioeconomic backgrounds?

The responses to this question, summarized in Table 6a on page 13, were disappointing. Sixty-nine of the 252 school districts reporting grouping either indicated that there was no information available regarding the number of students of low socioeconomic background or status (SES) involved in grouping in their districts or failed to respond to the question. Since the question was purposely asked in such a way that respondents

Table 6a
Percent of Homogeneously Grouped Students Who Are from Low Socioeconomic Backgrounds

	Northeast	South	Middle West	Southwest	West	Total
Less than 10%	20	6	13	2	11	52
10-25%	28	10	17	5	14	74
26-50%	11	14	8	7	4	44
51-75%	3	4	1	0	2*	10
More than 75%	1	0	2**	0	0	3
Information Incomplete***	13	13	20	8	20	74
No Grouping	12	18	9	7	20	66
Not Able to Respond	1	0	2	0	2	5
Total Number of Districts	89	65	72	29	73	328
Total Number of Students Involved	682,305	84,002	80,152	14,354	15,063 +	875,876 +

*The number of students involved in grouping was not reported.

**One school reported that 100 percent of its students moving from kindergarten to first grade were grouped but only a single class was involved.

***Includes 5 whose response to question 1 was recorded "Generally No, Unclassifiable."

to the questionnaire would not need to reveal information about the percent of students assigned to different groups who were of low SES, it is hard to believe that the high degree of unresponsiveness was by design. Still, approximate percents of low SES students involved in grouping should have been fairly easy to figure.

Table 6b, below, gives the approximate numbers of students involved in each of the categories reported by district in Table 6a.

Table 6b
Numbers of Low SES Students in Categories Shown in Table 6a

	Northeast	South	Middle West	Southwest	West	Total
Less than 10%	1,624	11,130	1,085	165	2,470	16,474
10-25%	43,698	20,978	6,867	2,637	8,642	82,822
26-50%	8,001	35,894	11,400	11,552	3,951	70,798
51-75%	508,482	16,000	10,800	000	?	535,282 +
More than 75%	120,500	000	50,000	000	000	170,000
Total Number of Students Involved	682,305	84,002	80,152	14,354	15,063 +	875,876 +

Question 7

What percent of these students are non-white?

For this question, too, the responses were disappointing. As shown in Table 7a on page 14, 56 of the 252 school districts reporting homogeneous grouping either indicated that information was not available concerning the racial composition of students involved in grouping in their district or failed to answer this question. Again, the question was purposely asked in such a way that respondents to the questionnaire would not need to reveal information about the percent of non-white students assigned to different groups. How-

ever, 22 percent of the respondents could not or would not answer the question as presented.

One observation is of special interest here. Forty-nine percent of the school districts in the Northeast and in the Middle West practicing ability grouping reported that fewer than 10 percent of the students involved were non-white; 29 of the 35 districts in the Middle West so reporting indicated that the percent of non-white students involved was less than one percent or zero. Many of the districts reporting low percents of non-whites in their grouping plans, particularly smaller districts in New England and in the Plains States, reported total non-white populations of less than one percent or zero by way of explanation of the absence of non-whites in their school populations and, hence, in their grouping plans.

Table 7a
Percent of Homogeneously Grouped Students Who Are Non-White

	Northeast	South	Middle West	Southwest	West	Total
Less than 10%	44	8	35	6	28	121
10-25%	11	16	7	7	3	44
26-50%	3	8	1	1	5	18
51-75%	4	5	0	1	2	12
More than 75%	0	0	1	0	0	1
Information Incomplete*	14	10	17	7	13	61
No Grouping	12	18	9	7	20	66
Not Able to Respond	1	0	2	0	2	5
Total Number of Districts	89	65	72	29	73	328

*Includes 5 whose response to question 1 was recorded "Generally No, Unclassifiable."

Table 7b, below, gives the approximate numbers of students involved in each of the categories reported by district in Table 7a, above.

Table 7b
Numbers of Non-White Students in Categories Shown in Table 7a

	Northeast	South	Middle West	Southwest	West	Total
Less than 10%	3,939	8,240	1,511	883	2,159	16,732
10-25%	6,288	35,600	7,650	4,442	414	54,394
26-50%	5,891	15,474	8,000	6,000	20,600	55,965
51-75%	545,842	3,793	000	150	25,000	574,785
More than 75%	000	000	287,736	000	000	287,736
Total Number of Students Involved	561,960	63,107	304,897	11,475	48,173	989,612

Question 8

What do you consider to be the advantages of homogeneous grouping in your school district?

As indicated earlier, the responses to this question and to question 9 are grouped according to the extent to which the school districts responding are currently practicing homogeneous grouping. For each group the responses are listed in order of the frequency with which they were mentioned by respondents.

It was expected originally that there might be wide differences in the nature of the responses given by the various groups since the questions asked specifically for "the advantages (and the disadvantages) of homogeneous grouping in your school district." Actually, the advantages and disadvantages listed for the different groups are very similar, except that the number of advantages and disadvantages bears a direct relationship to the extent to which homogeneous grouping is practiced.

Districts employing homogeneous grouping generally (180)

Improves attention to individual needs (45)
Permits students to progress at their own learning rate (36)
Allows the student to compete on a more equitable basis (33)
Reduces ability and achievement range within the classroom (25)
Facilitates curriculum planning (23)
Permits both remedial and enrichment programs (21)
Results in better teaching and more effective learning (18)
Makes it possible for each student to achieve success (18)
Permits the more effective selection and use of materials (17)
Makes instruction easier (13)
Reduces student frustration and dropout rate (10)
Is preferred by the teachers (8)
Improves teacher and student morale (6)
Encourages better use of teacher preparation time (5)
Permits more effective classroom planning (5)
Makes possible the development of advanced courses, sometimes with state aid, for the academically talented (5)
Offers no obvious advantages (4)
Reduces concentration on teaching average group (3)
Facilitates scheduling (3)
Improves the student's self-image (3)
Facilitates motivation (3)
Is liked by parents of more talented students (2)

Districts employing homogeneous grouping at some grade levels or in some subjects (35)

Makes it easier to adjust the curriculum to different needs and abilities (21)
Makes possible more economic and more effective use of materials and media (13)
Offers no obvious advantages (13)
Permits individual student to move at his own rate (10)
Offers every student an opportunity to achieve some success in school and to enjoy its attendant benefits—enhanced self-concepts, increased satisfaction with school, improved motivation to learn, and more rapid progress in learning (7)
Results in more effective teaching with fewer demands on the teacher (6)
Results in improved teacher morale (6)
Results in more time devoted to slow learners and consequent greater student involvement (4)
Simplifies scheduling procedures for the administrator (3)
Reduces teaching for the "middle" group (3)
Makes it possible to present esoteric concepts in accelerated classes that could not be presented in heterogeneous classes (3)
Decreases discipline problems and number of dropouts (2)
Permits students to move at their own rates in the basic skill areas at the same time allowing them the advantages of heterogeneous grouping in other subject areas (2)

Districts in which policies regarding homogeneous grouping vary from school to school (37)

Enables the teacher to work within the framework of one major lesson plan which can accommodate for student individual differences rather than many specific, diversified plans which may lead to teacher confusion and classroom chaos (13)
Permits more attention to individual student interests and problems (9)
Allows for enrichment, faster movement, and early graduation for the academically talented (7)
Permits the more efficient purchase and use of materials (3)
Makes it easier to stimulate motivation and, consequently, to improve class achievement (3)
Permits more attention to slow learners (3)
Motivates students to make better progress when in class of peers (2)
Provides better climate for instruction (2)
Reduces failure and retention (2)
Offers social advantages such as peer acceptance (1)

Reduces teaching for the "middle" group (1)
Improves administrative management (1)

Districts in which there is little or no grouping (71)

May offer better learning opportunities for students of other than average ability (6)
Pleases teachers who prefer this kind of organization (4)
Permits more concentration on needs of the individual student (2)
Improves the student's sense of accomplishment (2)
May be advantageous if groupings are flexible ones set up for specific purposes (2)
Permits better use of teaching aids (1)
Offers no obvious advantages (1)

Question 9

What do you consider to be the disadvantages of homogeneous grouping in your school district?

Districts employing homogeneous grouping generally (180)

Reduces or eliminates leadership and stimulation provided by heterogeneous grouping (37)
Stifles the socialization process, giving rise to snobbery in some cases and second class citizenry in others (30)
Fosters unhealthy self-concepts, especially among slow learners (24)
Results in labeling and stigma for slow learners (18)
Encourages some teachers to work under the misconception that since the class has been grouped according to ability, all students within that class are the same (17)
Destroys the spectrum of types with whom an individual functions in a real life situation (16)
Has no obvious disadvantages (15)
May result in separation of students by race and socioeconomic status (13)
Reduces attention to individual problems (12)
May create administrative problems, like arranging schedules (11)
Does not necessarily result in better learning (9)
Creates problems of parental understanding of potential of students at all levels (8)
Creates morale problems for teachers assigned to low groups (8)
Results sometimes in putting too many discipline problems together (5)
Is frequently based on invalid criteria (5)
Results in the formation of cliques (4)
Destroys the challenge of competition (4)
May lead to mediocrity in education (4)
Results in lowest level students getting least experienced teachers (4)
Denies enrichment programs for the brighter student (3)
Tends to "lock" slower learners (3)
Creates problems of student placement (3)
Results in inappropriate use of materials (2)
Creates social pressures (2)
Reduces flexibility (2)
Encourages dropouts (1)
Results in competition rather than cooperation (1)
Prevents bright students from becoming sensitive to problems of slow learners (1)

Districts employing homogeneous grouping at some grade levels or in some subject areas (35)

Tends to create a built-in expectancy for students to function at whatever level they are placed (16)
Denies the average and slow learner the stimulation of the more capable learner (12)
Provides a poor social-cultural mix (10)
Allows students little opportunity for movement throughout school years as a result of initial labeling (9)
Has no obvious disadvantages (8)
Results in parental objections on the basis of possible stigma (7)
Does not provide for individual needs (6)
Creates problems of leadership for the slower learner (6)
Tends to promote the idea of an intellectual elite, which is more status conscious and less tolerant (4)
Results in decreased motivation at all levels (3)
Damages the student's self-concept (3)
Results in assignment of reluctant teachers to slower classes (3)
Requires more effort to organize and schedule (2)
Is frequently based on invalid criteria (2)
Puts more discipline problems together (2)
Does not allow flexible grouping patterns in classroom (1)
Creates a situation that is not true to life (1)
Sometimes results in parental pressure for assignment to classes too advanced for the student (1)

Districts in which policy regarding grouping varies from school to school (37)

Creates a blighted teaching situation for the teachers of the slow groups (6)
Is likely to result in labeling and stigma (4)
Encourages tendency to ignore individual needs and consider all students alike (4)
Reduces opportunities for brighter students to stimulate the slower ones and for brighter stu-

dents to get ego enhancement from comparison with slower ones (4)
Creates problems of scheduling in the secondary school (3)
May set false standard that becomes self-fulfilled for some (3)
Tends to segregate students by race and socio-economic status (2)
Creates a situation that is not true to real life (2)
Does not provide a good social mix (2)
Does not inspire slower students (2)
Results in feelings of inferiority (2)
Does not adequately distribute leadership of students (1)
May result in development of cliques (1)
May result in lack of understanding of slower students by faster ones (1)
Creates too much feeling of self-importance in higher groups (1)
Tends to be too structured and rigid (1)
Causes difficulties because of wide age range (1)
Concentrates discipline problems (1)
Has no obvious disadvantages (1)

Districts in which there is little or no grouping (71)

Results in labeling, thus creating poor self-image for the slow and disadvantaged (10)
Reduces teacher and student enthusiasm and motivation (10)
Implies that class membership is determined by a constant set of factors with result that students, once grouped, will remain in those groups for a complete program (5)
Denies students the advantages of associating with others of different levels and abilities (5)
Tends to group students who are slow in one subject matter area in slow groups in all areas (4)
Denies slow students the leadership provided by higher groups (4)
Offers the slow learner little stimulation to succeed (3)
Results in segregation—racial, social, economic (3)
Has not been shown to improve learning—and may impede progress as the student progresses to higher grades (3)
Concentrates problems—both disciplinary and learning (2)
Impractical in schools with small enrollments or geographic problems (2)
Fosters antisocial attitudes that are not offset by any resulting gain from homogeneous grouping (2)
Limits class contact of talented students to other talented students, with consequent clashes of temper (1)
Creates a separation that is contrary to that of the world in which the child must function (1)

As indicated earlier, only two of the school districts responding reported that they are moving from heterogeneous toward homogeneous grouping. A number of districts, however, reported that while they are currently practicing homogeneous grouping to a considerable extent, the thrust is in the direction of heterogeneous grouping. A few comments from these districts follow.

In response to question 8 on the advantages of homogeneous grouping:

> At one time it was felt that by narrowing the achievement span, teachers could plan for more effective instructional experiences and that the learning patterns of students could be more scientifically utilized. Present emphasis upon individualized instruction is rapidly rendering this kind of thinking obsolescent in our district.

> Since our concept of grouping is one of ability grouping within subject matter, we believe the advantages are obvious. We think you should know, however, that in some subject areas we deliberately have heterogeneous grouping.

In response to question 9 on the disadvantages of homogeneous grouping:

> One disadvantage of homogeneous grouping is the step-ladder effect. In large schools with 20 to 25 sections to a grade, the achievement and ability levels of groups can become so unproductive that both teachers and students are constantly frustrated. Neither teachers nor students have the experiential background to cope with problems that arise.

> There are many effective arguments for strictly heterogeneous grouping and we are coming to this more and more.

The responses to question 8, generally, indicate that despite the fact that research on homogeneous grouping has failed to show that this practice results in significant increments in learning, school districts employing it can see advantages in their own situations and that even those districts not employing it can, nevertheless, name some advantages. The responses to question 9 show that districts employing homogeneous grouping are about as well aware of its disadvantages, either generally or in their own districts, as are those districts not employing it. In the face of the conflicting evidence offered by research and with the disadvantages that are obvious to the districts themselves, why does the practice of homogeneous grouping persist to the extent that it does?

One reason why homogeneous grouping is practiced widely is undoubtedly teacher preference for it. In a poll conducted by the NEA in 1961, a nationwide sample of public school teachers was asked the following question:

Considering all the advantages and disadvantages of ability grouping according to IQ or achievement scores, do you favor such grouping into separate classes . . . ?

Here are the answers received.

	Elementary	Secondary
Approve	57.6%	87.3%
Disapprove	33.1%	8.6%
Don't Know	9.3%	4.1%

Opinions were analyzed according to whether the teachers had or had not taught in schools with ability grouping. Elementary teachers who had taught under both arrangements were two to one in favor of ability grouping; and better than 90 percent of the secondary teachers who had taught under both arrangements were in favor of ability grouping.

In 1968 the NEA conducted a second poll on ability grouping. A scientifically selected sample of a nation's public school teachers was asked this question:

What types of pupils would you prefer to teach, so far as ability is concerned?

Four types of groups were listed: high, average, low, and mixed. In addition, respondents were allowed to indicate no preference. The results are shown below.

	Elementary	Secondary	Total
High	18.4%	34.6%	26.0%
Average	44.7%	38.9%	42.1%
Low	4.3%	1.9%	3.1%
Mixed	21.3%	15.2%	18.4%
No Preference	11.3%	9.4%	10.4%

It is interesting to note that more teachers prefer to teach classes of average ability than classes of any other type. And, as one might expect, with an overwhelming number of teachers expressing preferences, only 3 percent prefer to teach classes of low ability. As to grade levels, the elementary teachers choose mixed and high groups only half as often as average groups, with a slight preference for mixed over high groups. The secondary school teachers prefer high groups almost as much as average groups, while mixed groups run a poor third.

SUMMARY AND CONCLUDING REMARKS

The information assembled permits several generalizations. Briefly, if the school districts sampled are in any way representative, it may be said on the basis of responses to the questionnaire that:

1. Ability grouping is being practiced in some form in approximately 77 percent of the nation's public schools.
2. There is proportionately more grouping in the Northeast and the Middle West than in other parts of the country.
3. Slightly more than 20 percent of the schools use grouping at all grade levels, with more grouping being done at the secondary school level than at the elementary school level.
4. Only about 22 percent of the schools practicing grouping have been doing this for 16 years or more.
5. Tests are used by about 82 percent of the schools that practice grouping, but only about 13 percent among these rely on test scores alone; rather, they use them as one of two or more criteria for grouping.
6. The larger the school district, the more likely it is that grouping will be practiced on a systemwide basis.
7. About 23 percent of the students involved in grouping are "known" to be from low socioeconomic backgrounds.
8. About 26 percent of the students involved in grouping are non-white.
9. In school districts where grouping is employed, it is favored more often than not because it is seen as a convenient way to provide for individual differences, to make teaching easier, and to facilitate curriculum planning.
10. In school districts where grouping is not employed, it is seen as likely to result in the labeling of students too early in their school careers, to limit the possibilities of movement of students with maturation, and to reduce both teacher and student motivation.

It must be repeated that the failure of many school districts to respond to certain questions in the questionnaire may have implications for the study and render some of these generalizations erroneous.

REFERENCES

Coleman, James S. and Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. *Equality of Educational Opportunity.* Washington, D. C.: Government Printing Office, 1966.

Gore, Lillian L. and Rose E. Koury. *A Survey of Early Elementary Education in Public Schools, 1960-61.* U.S. Department of Health, Education, and Welfare, Office of Education, Circular No. 782. Washington, D.C.: Government Printing Office, 1965.

National Education Association, Research Division. Research Memo 1962-29, *Ability Grouping,* Washington, D. C.: National Education Association, September 1962.

National Education Association, Research Division. Research Summary 1968-S3, *Ability Grouping,* Washington, D. C.: National Education Association, 1968.

National Education Association, Research Division. "Ability Grouping: Teacher Opinion Poll." *NEA Journal,* 57:53, February 1968.

National Education Association, Research Division. "Teacher Opinion Poll," *NEA Journal,* 50:62, April 1961.

PRINTED REPORTS SUBMITTED BY SCHOOL DISTRICTS

Coffin, Gregory C. and Laval S. Wilson and Annette Grubman. *Grouping Children in Integrated Schools,* Resource Manual. Evanston, Illinois: Community Consolidated Schools, 1968-69.

Cohen, Ruth Eisenberg. *Grouping for Instruction,* Research Study Series 1966-67, Research Report Number 1. Los Angeles, California: Los Angeles County Superintendent of Schools Office, Division of Research and Personnel Services, June 1967.

Counce, D. Shelby. *The Memphis Story—Mobilizing a Large School System and a Community for the Superior and Talented School Project.* Chicago, Illinois: STS Project, 1965.

Curry, Robert P. et al. *Marking and Reporting Practices.* Cincinnati, Ohio: Cincinnati Public Schools, 1968.

Eisenstadt, Thomas S. et al. *Individual Progress Program.* Boston, Mass.: The School Community of the City of Boston, 1968.

Erbling, Harold, H. et al. *Grouping for Quality Learning.* Columbus, Ohio: Columbus Public Schools, 1969.

Joselyn, Edwin Gary, Editor. *A Study of Testing Practices in Minnesota Public Schools.* St. Paul, Minn.: Minnesota State Department of Education, 1967.

Los Angeles County Superintendent of Schools Office. *Guiding Today's Children: A Guidance Book for Teachers and Administrators of Elementary Schools.* Los Angeles, California: California Test Bureau, 1959.

Los Angeles County Superintendent of Schools Office. *Guiding Today's Youth: A Guidance Book for Teachers and Administrators of Secondary Schools.* Los Angeles, California: California Test Bureau, 1962.

Ohrenberger, William H. and Mary E. Vaughan, Marion J. Fahey, et al. *Guide Book for Teachers of Advanced Work Classes, Grades V and VI.* Boston, Massachusetts: Boston Public Schools.

THE QUESTIONNAIRE

UNIVERSITY OF GEORGIA

COLLEGE OF EDUCATION
ATHENS, GEORGIA 30601

QUESTIONNAIRE ON SCHOOL GROUPING PRACTICES

1. Are students at any grade level in your school district grouped homogeneously?
2. If so, at what grade levels is homogeneous grouping done?
3. How long has homogeneous grouping been practiced in your district?
4. On what basis are your students assigned to homogeneous grouping? (If on the basis of test scores, please name the test.)
5. How many students in all are involved in your homogeneous grouping plan?
6. What percent of these students are from low socioeconomic background?
7. What percent of these students are non-white?
8. What do you consider to be the advantages of homogeneous grouping in your district?
9. What, if any, do you consider to be the disadvantages of homogeneous grouping in your school district?

II. THE IMPACT OF ABILITY GROUPING ON SCHOOL ACHIEVEMENT, AFFECTIVE DEVELOPMENT, ETHNIC SEPARATION, AND SOCIOECONOMIC SEPARATION

OVERVIEW

The quality of an educational environment may be defined as the quality of the experiences that are provided by that environment. Thus, the extent to which ability grouping tends to enhance or reduce school learning experience is of particular educational significance. If ability grouping tends to restrict the quality of children's school experiences, such practices by design, if not intent, foster an unsound environment for the education of children and should be discontinued. If, on the other hand, evidence suggests that ability grouping tends to maximize the cognitive and social experiences available in a classroom, then such practices should be initiated and/or continued in the interest of maintaining quality education.

Ability grouping is the practice of organizing classroom groups in a graded school to put together children of a given age or grade who have most nearly the same learning achievement or capability, largely on the basis of standardized tests. In the survey conducted as part of the present study, 206 of the 252 school districts reporting the use of ability grouping, or 82 percent, use standardized tests as an integral feature of the process. (See Table 4 in Section I.) In the discussion that follows, all such standardized tests, whether of subject matter achievement, IQ, or "aptitude," are considered simply different varieties of achievement tests. This terminology is intended to reflect that, functionally, the usual distinction between measures of aptitude and achievement, that is, innate talents vs. learned talents, is not a meaningful and worthwhile division. In classifying IQ and other aptitude tests, as well as reading, arithmetic, and other subject matter tests, as measures of achievement, the implication is that a score obtained on each of these instruments reflects the child's level of knowledge in a given subject or skill which, in turn, reflects an environmental and/or developmental end product at a specific point in time.

There are a number of dimensions on which one may evaluate the quality of a particular educational environment. Chief among such dimensions is student achievement in the basic academic skills, that is, reading and arithmetic. For more than five decades, educators and researchers have focused on these dimensions and have contributed a large body of relevant data. Recently, a second dimension has received research attention. This dimension can be broadly classified as social learning. Here student attitudes and aspirations, personality development, adjustment to school, social behaviors, and so forth, are measured to determine in what ways heterogeneous and homogeneous grouping practices influence such affective development. Few research efforts have at any time been directed at a third dimension, the practical consequences for ethnic and socioeconomic separation of an ability grouping policy. These are consequences that heretofore have not been considered important to the academic and social growth of children.

It is not the purpose here to present a detailed review of this research but rather a digest of the research literature which has led to our findings, namely, that grouping practices based on standardized measures of achievement not only tend to restrict the quality of the instructional experiences of children with respect to academic and social learning, but also, as a result of ethnic and socioeconomic separation, tend to restrict the overall range of experiences and learning opportunities available in the classroom.

DEFINITIONS AND DISCUSSION*

In public education, the term "grouping" has been a broad rubric subsuming a wide variety of organizational plans, selection criteria, instructional methodology, and educational philosophies. Since the school has traditionally been defined by its group setting, methods have had to be devised to make the instruction of groups of children more effective and/or more manageable. The major options for vertical organization have been graded, multigraded, or nongraded (continuous progress) schools. Whichever of these plans exists in a school, a concomitant pattern of horizontal organization, which assigns students to classes, teachers, rooms, and curricular programs, must emerge.

Homogeneous grouping occurs when classes are formed on the basis of similarity on some specific characteristic of the students. The criterion for this classification may be age, sex, social maturity, intelligence, achievement, learning style, or a combination of these. The group, however, is homogeneous only with respect to this one criterion, or combina-

*This part relies heavily on a paper prepared for Dr. Edmund W. Gordon, Director, ERIC Information Retrieval Center on the Disadvantaged, Teachers College, Columbia University, by S. Bernstein and D. Esposito, *On Grouping in the Experimental Elementary School Project,* November 1969.

tion of criteria. In practice, of course, it is impossible to form a group of individuals possessing the identical degree of any characteristic other than sex or other nominal variable like skin pigmentation or eye or hair color, so the objective for homogeneity is to produce a reduced range of a particular characteristic in the group. Ability grouping is one of the many forms of homogeneous grouping, and generally refers to the use of standardized measures of intelligence, ability, or achievement in a given subject in classifying students into separate ability categories.

When ability grouping is applied to all grades and used throughout a school system, it is usually called "tracking." In secondary schools, children are assigned to clearly labeled curricular tracks, that is, college preparatory, vocational, commercial, general, or technical. Practically, this means that for ninth-grade mathematics, a student will be assigned to algebra, business mathematics, or basic mathematics, depending on the track in which he is enrolled. Similarly, students enrolled in the college preparatory track may be exposed to biology, chemistry, and physics, while vocational or general students are limited to general science and biology. In addition, students are often further channeled into biology for college preparatory enrollees and biology for general or vocational enrollees. In short, ability and track-type arrangements tend to divide and separate students for instructional purposes. At the elementary school level, this results in a reduction in the frequency, range, and quality of contacts that a student has open to him; at the secondary school level, it further means that a student is enrolled in a set program that leads to a set destination or diploma at the end.

On the other hand, if one is concerned with achieving a mixture of children in a given classroom who differ on a number of dimensions, including "ability," a heterogeneous grouping policy can meet this concern. Heterogeneous grouping is generally accomplished by assigning children to classes alphabetically or by choosing every *n*th name on a list. Less often, classes are deliberately structured so that a wide range of ages, abilities, achievement levels, socioeconomic backgrounds, and ethnic status is assured in each class.

Homogeneous and heterogeneous grouping concepts are essentially at opposite ends of the same continuum. Inasmuch as homogeneous grouping can theoretically occur only with respect to nominal variables, it seems evident that homogeneous grouping serves merely to restrict the range of individual differences with respect to certain continuous or ordinal criterion dimensions, while heterogeneous grouping tends to expand the range of individual differences on all dimensions. It is impossible to achieve truly homogeneous grouping, even along a single variable, since test data and other measures used in ability grouping are not generally reliable enough for such categorizing. Homogeneous grouping may merely result in less sensitivity to individual differences in children by giving teachers the false notion that students in these classes are almost identical in achievement, learning style, and social needs, that is, that the different patterns of abilities that they expect to emerge in heterogeneous groups will not emerge in homogeneous groups.

Clark (1963) has cautioned: "Probably the chief argument against homogeneous grouping is the fact that children so segregated lose their individuality in the educational situation.... Homogeneous groupings tend to require that children be seen in terms of group characteristics rather than in terms of their individual characteristics."

What little research has been done with respect to the ethnic and socioeconomic effects of homogeneous grouping shows that such grouping tends to segregate along ethnic and socioeconomic lines as well as on ability, probably even more sharply. In commenting on this point, Passow (1967) observed that some educators would argue that

> ... ability grouping is simply a means of making respectable the procedures whereby pupils from lower socioeconomic and racial or ethnic minority groups are relegated to the "slower" and "nonacademic" programs and provided with a basically inferior education. Observers of racially mixed schools frequently find that ability grouping is a means by which pupils are re-segregated within the school.

The criteria for grouping students in studies which examine the effects of ability grouping have, more often than not, been measures of "intelligence" or of achievement, ranging from several different measures of reading achievement to scores on a single arithmetic subtest of an achievement battery. Grouping on the basis of scores on IQ tests assumes that mental age and ability are synonymous as well as that a uniform level of abilities characterizes each individual. Reading and arithmetic tests may not measure functional verbal or mathematical ability or take into account the variety of factors that influence an individual's test score. Particularly with young children, it is doubtful that any of these measures are accurate or valid for grouping.

Table 8 shows how differently children in an ordinary seventh-grade population of 103 achieve in the basic subjects of reading and arithmetic. Barely half of the group, 55, would be classified in the same third of the total group on both measures. Note that six stand in the top third in one subject and the bottom third in the other.

Table 8

Scatter-Diagram of Achievement Scores in Reading and Arithmetic, Stanford Achievement Test, Advanced, Form J, for 103 Pupils in Grade VII in a School with Average Achievement

Average Reading \ Average Arithmetic	4.5–4.9	5.0–5.4	5.5–5.9	6.0–6.4	6.5–6.9	7.0–7.4	7.5–7.9	8.0–8.4	8.5–8.9	9.0–9.4	9.5–9.9	
11.0–11.4								2	1			
10.5–10.9							2					
10.0–10.4					1		3	1	1	1		
9.5–9.9							1	2				36
9.0–9.4						1	1	2	3	1	1	
8.5–8.9				1		4	1	1	3	2		
8.0–8.4				1	2	3		1		1		
7.5–7.9			2		2	3	3	2	2	1		35
7.0–7.4					3	2	5	2				
6.5–6.9			1	2	1	2	2				1	
6.0–6.4	1		1	2	3		1	2	1			
5.5–5.9				2	1	4	1					
5.0–5.4												32
4.5–4.9		1	1									
4.0–4.4			1									
3.5–3.9					1							
			30			39		34				

The dependent measures employed in studies of ability grouping present further problems. Most studies examine the effects of grouping practices on academic achievement measured by standardized tests. Some use measures of attitude and personality development, social learnings, adjustment to school, or teacher reaction. Only a few, however, have used a multivariate approach to examine differential effects of ability grouping along a number of dimensions. Hence, rarely have the arguments for or against homogeneous grouping listed in Section I been tested empirically.

The major purpose of reducing the range of ability in any classroom is, ostensibly, to provide more easily for individual differences. Research studies rarely specify, however, the ways in which instruction has been adapted or modified from group to group. It is generally implied that the curricular programs, the methodology, and/or the pace have been varied. Yet, there appear to be no studies which measure instructional practices to show whether those practices have been kept constant or varied over experimental and control groups.

Goldberg et al. (1966) summarize some of the many difficulties of interpreting research in ability grouping. They point out that studies vary considerably in their range of objectives, in the basis for determining "homogeneity," in duration, in adequacy of selection bases and means of matching experimental and control groups, in numbers of students, numbers of groups, and size of classes, in differentiation of curricula and teaching method, in instruments and techniques used in assessing changes in students, and in the preparation of teachers for various groups, and have generally failed to examine effects of grouping on teachers and administrators.

If it is assumed that the variables indicated above, either independently or in combination, affect student achievement, then not controlling for these variables in studies of ability grouping tends to minimize the variation between or among ability groups, thereby tending to reduce the likelihood of finding statistically reliable differences. With this perspective, then, it is not surprising to find that research results are inconclusive. No clear and consistent effects on aca-

demic achievement have been found. Efforts on students' attitudes towards themselves and towards school are also ambiguous.

Eash in a 1961 summary of ability grouping research offers several conclusions that speak to some of the major issues related to homogeneous and heterogeneous grouping practices. These conclusions are:

1. Ability grouping in itself does not produce improved achievement in children.

2. Contrary to statements in previous summaries of the research on the effects of ability grouping on children's achievement..., more recent research evidence seems to indicate that ability grouping actually may be detrimental to children in the average and lower ability groups.

3. Ability grouping at an early age seems to favor unduly the placement of children from the higher socioeconomic class in higher ability groups.

4. Research evidence in the area is quite meager, but what is available does not support the prevalent assumption that college achievement is improved by abiiity grouping.

5. Ability grouping as an organizational structure may accentuate the attainment of goals, and symbols for goals, of narrow academic achievement to the extent that other broader desirable behavioral goals and objectives are attenuated and jeopardized.

6. The evidence is fairly conclusive that grouping practices in a school can assist in developing social situations that influence the student's perception of self, his sense of dignity and worth, and his attitudes toward other children. In view of this, grouping practices should be concerned with furthering the establishment of social climates that will encourage the intellectual, social, and personal development of every child without detrimental effects on individual children.

7. Grouping practices are significant factors in establishing a teaching-learning situation whereby children can acquire the general education skills and abilities needed by all citizens in a democratic society. This means, in brief, that students need opportunities to work in common purpose with a wide range of individuals. Grouping practices which separate students on the basis of ability as by group IQ or standardized achievement tests reduce the likelihood that students will be exposed to a broader range of ethnic and cultural differences in the society.

8. Pressures to institute certain grouping practices in our schools represent pervasive social problems in our culture. Educators need to be doubly alert that the schools are not utilizing grouping practices which assist in maintaining and promoting social and racial biases which militate against the general education objectives, equal educational opportunity, and the development of each person as an individual.

If the major educational objective of classifying children into restricted range classroom environments is "greater provision for individual differences," and since there is no clear-cut evidence indicating that this objective has been realized in the tens of thousands of homogeneous classrooms across the nation, then one is compelled to conclude that ability grouping, as presently implemented, has failed to establish its merit as a sound educational policy. In this, we second the conclusion put forth in NEA Research Summary 1968-S3:

> Despite its increasing popularity, there is notable lack of empirical evidence to support the use of ability grouping as an instructional arrangement in the public schools.

The logical implication of these findings is to engineer an educational environment that can practically sustain learning task-oriented small group activities in which more direct individual attention and instruction can be realized.

GROUPING PRACTICES AND SCHOOL ACHIEVEMENT

The literature of better than sixty years relating to research on grouping practices and school achievement has been systematically and thoroughly reviewed by many individuals and groups. Probably the most comprehensive and authoritative reviews have been those of Billett (1932), Ekstrom (1959), Borg (1966), the Research Division of the National Education Association (1968), and three contributors to the *Encyclopedia of Educational Research:* Otto (1941, 1950), Goodlad (1960), and Heathers (1969).

Each of these reviews is accompanied by an extensive bibliography. Taken together, these bibliographies list hundreds of different studies of the relationship between grouping practices and school achievement. While many of the earlier studies, and some of the later ones as well, would not be considered today to be truly "research" studies, each of them has information to offer the individual who is interested in pursuing a study of grouping from the beginning.

Billett (1932) reviewed 140 research studies made between 1910 and 1928. He classified 108 of these as "experimental or practical." Of the 108 studies, how-

ever. Billett listed only four as "thoroughly controlled"* and two as "partly controlled." Of the four "thoroughly controlled" studies, two were favorable to grouping, one was doubtful, and one was unfavorable. One of the two "partly controlled" studies was favorable to grouping and the other was unfavorable.

Otto (1941) summarized the status of ability grouping as of that date. His conclusions may be summarized as follows: (1) Where adaptations of standards, materials, and methods had been made, the evidence slightly favored ability grouping as contrasted with heterogeneous grouping. However, (2) the evidence of the relative merits of various adaptations of standards, materials, and methods was too inadequate to form a judgment. (3) The greatest relative effectiveness of ability grouping appeared to be for "dull" children, the next greatest for average children, and the least (frequently harmful) for bright children. (4) Evidence regarding particular grade levels or subjects in which ability grouping was especially effective was too inadequate to form a judgment. (5) Most teachers preferred to work with homogeneous rather than with heterogeneous groups. (6) On the whole, parents were favorably disposed to the use of grouping. (7) Although one study showed the great majority of students in schools using ability grouping to be satisfied and happy, evidence regarding the effect of ability grouping on characteristics of students other than knowledges and skills was highly subjective and inconclusive. (8) In general—and this is perhaps Otto's most important conclusion—variability in achievement in ability groups was almost as great (74 to 93 percent under varying conditions) as it was in unselected groups.

Nine years later Otto (1950) reported that his search of the literature on ability grouping showed no research studies to have been made for 15 years. The conclusions reported by him at this time were, therefore, the same as those reported earlier.

Ekstrom (1959) reviewed 33 research studies made between 1923 and 1959. She found 13 studies, with differences having or approaching significance,† which favored homogeneous grouping; 15 studies reporting no differences in achievement between homogeneous and heterogeneous groups, or differences unfavorable to homogeneous grouping; and five studies reporting mixed results, partly favorable and partly unfavorable to homogeneous grouping. Ekstrom could find no consistent pattern for the effectiveness of homogeneous grouping related to age, ability level, curriculum, or method of instruction. She cautioned that the differences in number of favorable or unfavorable studies should not be considered too seriously since the studies differed so widely in quality, purpose, and scope. She noted the inability to control certain relevant factors like the type of teaching and the differentiation of teaching according to ability levels as important weaknesses in most of the studies. She was also critical of the experimental design in several of them, especially the use of matched pairs of subjects based on unwarranted assumptions of similarity in other respects.

Goodlad (1960), who reviewed 12 pieces of literature regarding ability grouping incidental to a review of classroom organization generally, reported conclusions in part reminiscent of those of Otto 19 years earlier: (1) Evidence with regard to academic achievement appeared to favor ability grouping slightly for slow students and to a greater extent for bright students. (2) The grouping itself was not so significant a contributor to academic achievement as was differentiation by curriculum. (3) Studies of ability grouping in different subject matter areas were somewhat contradictory. (4) Teachers reacted more favorably to teaching homogeneous groups than to teaching heterogeneous groups.

Borg (1966) reviewed 37 research studies made between 1922 and 1962, 20 of them being studies that had also been reviewed by Ekstrom. His findings confirmed the inconclusiveness found by earlier reviewers to be true of studies on grouping practices and school achievement made prior to the early 1960's. Of the 37 studies, Borg found 20 with differences of significance

*In a controlled study of the effects of ability grouping, the investigator provides evidence that the effects of other possible causes of differences between the groups being compared have been "controlled," that is, the groups have been matched on the possibly influential variables or statistical procedures have been applied to correct for the possible effects. In an uncontrolled study, the true cause remains in doubt.

†Here and hereafter in this document, the term "significance" will be used in its technical statistical meaning. That is, a difference in favor of one method of grouping or another will be pronounced "significant" if appropriate statistical checks indicate that so large a difference would arise as a matter of chance variation between random samples of the same sizes so infrequently that it is most reasonable to dismiss this possibility. Instead, it is better to presume that the difference found is attributable to factors that will cause differences in the same direction to occur whenever similar samples are compared that differ in respect to grouping. We speak of differences being "significant at the 5 percent level" when differences as large or larger would be expected to be found in less than 5 percent of pairs of random samples of these sizes drawn from a common pool of individuals who had been taught under identical circumstances. We speak with even more confidence of the "significance" of a difference if the likelihood of occurrence of one so large between random samples of these sizes from a common pool is less that 1 in 100; in that case, we speak of the difference as being "significant at the 1 percent level."

or approaching significance. Of these 20 studies, 13 were favorable to homogeneous grouping and seven were unfavorable.

NEA Research Summary 1968-S3, with 158 bibliographical entries, reports three reviews not covered earlier. Eash (1961) reviewed 28 items. His conclusions have been presented in detail earlier in this section (p. 23). Wilhelms and Westby-Gibson (1961) concluded that (1) there was no evidence that ability grouping per se was leading to improved mastery of subject matter; (2) the evidence slightly favored ability grouping, but the difference was small; (3) if any group had gained from ability grouping, it had been the low group rather than the ablest group; and (4) teachers tended to favor grouping as easing their problems of instruction, Franseth (1964) suggested that the findings reviewed by her raised as many questions as they answered. On the basis of her study, she concluded that factors other than grouping procedures might well account for differences in gains in achievement when they occurred between children homogeneously and heterogeneously grouped.

NEA Research Summary 1968-S3 also abstracted a total of 50 research studies on the effects of ability grouping published since 1960. Twenty-three of the studies were concerned with ability grouping at the elementary school level, that is, in grades 1 through 6; 23 were concerned with ability grouping at the secondary school level, that is, in grades 7 through 12; and four were concerned with ability grouping at both elementary and secondary school levels. Of the 50 abstracts, 42 pertain to the effects of ability grouping on academic achievement. From these 42 abstracts it is possible to infer again that, although the research on the effects of ability grouping on school achievement is extensive, the results, in general, are inconclusive and indefinite; and that factors other than ability grouping account for the differences in achievement that appear when learners grouped according to their abilities are compared with their counterparts in heterogeneous or randomly grouped situations. In this connection, where ability grouping appears to be more successful than heterogeneous grouping, modifications in educational objectives, curricular organization, teaching methodology, and teaching materials may well contribute more to the differences than does ability grouping itself. Some of the research studies abstracted in the NEA Research Summary are described in more detail later in this section.

Heathers (1969), with 84 bibliographical references covering the period from 1932 to 1968 but concentrating particularly on the literature of the 1960's, indicates that the major research studies reported in the 1960's lend strong support to the more recent view that ability grouping is associated with detrimental effects on slow learners, who, when they are placed in low ability groups, have been found to attain lower scores on achievement tests than comparable students obtain when taught in heterogeneous groups. One possible explanation for this phenomenon, Heathers notes, is that slow learners, in the absence of superior students, have fewer opportunities to learn vicariously through paying attention during classroom discussions in which they can be stimulated by other students. Another possible explanation is the self-fulfilling prophecy, that is, if teachers expect less from students who are assigned to low groups and teach them correspondingly less, the students who are assigned to such groups generally expect less of themselves and behave accordingly; on the other hand, when slow students are assigned to other groups, they are more successful. Heathers also reports evidence that the quality of instruction offered low groups tends to be inferior to that provided groups comprised of abler students. He reports that teachers have indicated that they tend to stress basic skills and factual information with slow learners and use drill with great frequency; conversely, they tend to stress higher levels of conceptual learning with high ability students and encourage them to conduct independent projects.

Heathers also mentions the assumption that ability grouping reduces the range of learning-related differences within a group, and that this reduction of range facilitates teaching and learning. This assumption, however, he explains, tends to be invalidated by the fact that the characteristics of students as learners are not adequately represented by their scores on general intelligence tests. A given student's ease and rate of learning and his level of achievement vary considerably from one curricular area to another, and from topic to topic and from task to task within each area. When students are grouped on the basis of intelligence quotients alone, the range of scores on achievement tests is still great.

Heathers suggests that the most effective way to reduce the range of a class in achievement would be to group differentially subject by subject and to base this grouping on separate measures of achievement for each area. He points out, however, that within such groups there would still remain large differences in ability and many other variables that influence learning.

Heathers also deals with the widely held notion that in ability groups rapid learners are freed from instruction which is geared to less capable students, and that since they are challenged to keep up with their intellectual peers, their achievement is enhanced. Related to this is the further notion that slow learners benefit from instruction geared to their capacities and from experiencing success more often in the absence of

abler students. Heathers indicates that these assumptions are of at least questionable validity. He reports evidence that placing a student in a group designated as low or slow stigmatizes the student, and that this is reflected in the student's losing interest in learning and study, thereby further debilitating his achievement.

A direct quotation from Heathers pretty well summarizes the inferences he derived from the evidence he found in the literature on ability grouping he reviewed:

> Writing an epitaph for grouping may well be the task of the reviewer of research on grouping for the 1980 edition of this encyclopedia [that is, the *Encyclopedia of Educational Research*]. Even today it appears that grouping as a central theme of organization for instruction has nearly run its course and is in the process of being replaced by a familiar theme—individualized instruction—that became a focus of educational reform in the mid-1960's.

Significant Research Studies of Achievement Effects From 1960 to the Present

As indicated earlier, NEA Research Summary 1968-S3 contains abstracts of 50 selected research studies on ability grouping which have been published since 1960. Forty-two of these are concerned in whole or in part with the effects of ability grouping on school achievement. The most significant studies will be reviewed again in some detail in this section. In addition, other significant studies not reviewed elsewhere will be reported.

The two most carefully designed and most rigorously controlled studies reported in NEA Research Summary 1968-S3 are those done by Borg (1966) and Goldberg et al. (1966). Both studies were longitudinal, the Borg study being conducted over a period of four years and the Goldberg study for a two-year period.

Borg (1966) used two adjacent and closely comparable school districts in Utah. In one district students were placed in ability groups on the basis of composite scores on an achievement test battery, and an attempt was made to adapt curricular materials to the different ability levels and to adjust the rate of presentation to the level of the individual students. In the other district a program of random grouping with enrichment, that is, an attempt to adjust the depth of learning to individual differences, was employed. In the first year over 2,500 students from grades 4, 6, 7, and 9 were selected for the study; during the second year the sample was increased to about 4,000 students.

In the Borg study, students tested in grade 4 were followed through grade 7; other grade samples were similarly followed over the four-year period of the study. Thus, data were collected from all grades from 4 through 12. The *California Achievement Tests* were used during the pilot study year; the *Sequential Tests of Educational Progress* were used during the final three years of the study.

Borg reported 54 statistical comparisons between randomly grouped and ability-grouped elementary school students. Of the 54, 28 were statistically significant at either the 5 percent or 1 percent level; 19 of the significant differences were found to be favorable to ability-grouped students, while nine favored randomly grouped students. However, since 15 of the 19 significant differences favoring ability-grouped students occurred during the first year of the study, the Hawthorne Effect* apparently operated rather strongly in favor of the ability groups during that year. If the first-year differences had been due primarily to the true superiority of ability grouping over random grouping, the differences would have increased each year as the cumulative effects of the more effective system widened the achievement gap between the two groups. This did not occur; in fact, most of the achievement differences which favored the ability-grouped students disappeared by the time these students had completed the sixth grade.

For elementary school students, Borg reported 18 achievement comparisons where superior students were the foci. Of the 18 comparisons, 11 were statistically significant, with 10 of these 11 favoring ability-grouped students. In terms of overall achievement differences for the four years of the study, ability-grouped superior students were significantly higher than randomly grouped superior students. For average students, however, Borg found no consistent trend favoring either random or ability grouping. In the comparisons between slow students, six significant differences were reported by Borg, with four of the six favoring the randomly grouped slow pupils. When the Hawthorne Effect, which operated on the ability-grouped students during the first year of the study. is taken into consideration, the relatively greater gains of the randomly grouped students are of even greater educational significance. Borg, in this connection, writes: "All in all, we may conclude that neither ability grouping with acceleration, nor random grouping with enrichment, is superior for all ability levels of elementary school pupils. In general, the relative achievement advantages of the two grouping systems were slight, but tended to favor ability grouping for superior pupils and random grouping for slow pupils.

*The Hawthorne Effect describes temporary gains that take place because of the novelty of the experimental treatment rather than permanent gains that may take place as a result of the treatment.

As was hypothesized, the differences for average pupils did not consistently favor either grouping treatment."

Since all five of Borg's samples were in junior high school sometime during the four years of his study, it is possible to draw inferences with respect to the relationship between ability grouping in the junior high school and achievement. When the achievement data for the five samples were combined, 60 statistical comparisons between comparable ability-grouped and randomly grouped students were made: 33 in mathematics and 27 in science. Of the mathematics comparisons, five were significant in favor of the ability-grouped students and five in favor of the randomly grouped students, while the other 23 were non-significant. Of the science comparisons, five significantly favored ability grouping and one significantly favored random grouping, while the remaining 21 were non-significant. When Borg's junior high school data were examined for superior, average, and low ability levels, there was a slight tendency for ability grouping to produce higher mathematics achievement among superior students and higher science achievement among average students. Among slow students, random grouping tended to produce higher achievement in both mathematics and science.

Of 30 comparisons made by Borg between achievement in mathematics and science for ability-grouped and randomly grouped students in senior high school, only four of the comparisons were significant. All four favored ability grouping, and all four differences were in mathematics achievement: one for superior students, two for average students, and one for slow students. It should be noted that less confidence can be placed in Borg's findings on the high school years than in the elementary and junior high school years because of the relatively small amount of high school data.

From his total data on ability grouping and school achievement, Borg found it possible to state the following conclusions: (1) At the elementary school level, the superior student generally showed greater gains in ability-grouped classes; for average students the pattern of advantages and disadvantages associated with the two grouping treatments was so complex that there was nothing to permit a choice between the two grouping treatments; the slow students generally showed better performance in the heterogeneous classrooms. (2) At the junior high school level, ability grouping led to significantly greater achievement gains for superior students although these differences were not large; for average groups the pattern was somewhat the same, with ability-grouped students making higher achievement scores; slow students in randomly grouped classrooms achieved more than their ability-grouped counterparts. Borg offered these conclusions, however, with the caution that they reflected his own value system and that educators having different orientations might well draw different overall conclusions from the findings of his research. Our conclusion is that his findings may be taken at face value, but with particular note of (1) the large proportion of comparisons (96 of 144) that failed to yield significant differences despite the large samples; (2) the failure of significant differences favoring homogeneous grouping at the end of the first year at the elementary school level to persist or increase thereafter; and (3) the fact that whatever modest significant differences favored homogeneous grouping were at the superior level, while low-ability level students tended to do somewhat better in heterogeneous groups.

The study of Goldberg et al. (1966) involved about 2,200 students in grades 5 and 6, organized into 15 grouping patterns in 86 classes in 45 New York City elementary schools. The grouping criterion was intelligence, and five ability levels were designated: (a) gifted, IQ 130 and over; (b) very bright, IQ 120-129; (c) bright, IQ 110-119; (d) average, 100-109, and (e) low or below average, IQ 99 and lower.

The authors set out to investigate three null hypotheses: (a) The presence or absence of extreme ability levels (gifted or slow) has no effect on the changes in performance of other ability levels. (b) Narrowing the ability range in the classroom has no effect on changes in the performance of students. (c) The relative position of any ability level within the range has no effect on changes in the performance of students. The hypotheses were tested for five major variables: (a) academic achievement, (b) self-concept, (c) interest and attitudes toward school, (d) assessment of more and less able peers, and (e) teacher ratings of students. Only the first of the variables will be discussed here; the others will be discussed later in this section.

In general, the results showed that in predominantly middle-class elementary schools, narrowing the ability range in the classroom on the basis of some measure of general academic aptitude will by itself produce little positive effect on the academic achievement of students of any ability level. In contrast, presence of gifted students in a class tends to raise science achievement of all levels of students, while presence of low ability students has a similar positive effect on arithmetic achievement.

Assessment of the various ranges of grouping patterns showed the broadest pattern to be generally somewhat more effective than any of the combinations of patterns with narrower ranges. A most significant finding was that gains in achievement were more strongly influenced by teacher differences and

group differences in individual classrooms than by the presence or absence of high ability students, the range of ability in the class, or the intellectual ability of the students. Between-class variability was greatest for the gifted students and least for the slowest students. When teacher effectiveness across ability levels was analyzed, it was found that teachers were more effective in teaching one or two subjects to a wide-range ability group than in teaching several subjects to a narrow-range ability group. In fact, most teachers were more effective in teaching one subject to several ability groups simultaneously than in teaching all subjects even in narrow-range classes. Finally, average achievement across all subjects was greatest in classes including four or all five of the ability levels described earlier in this summary.

Locke (1962) studied the effect of separating rapid learners from non-rapid learners for instruction in the intermediate grades. Criteria for determining rapid learners included scores above the 89th percentile on the *Otis Quick Scoring Test of Mental Ability* and the *Iowa Tests of Basic Skills* and consistently high school marks in grades 3 and 4. In the experimental group, rapid learners were homogeneously grouped in one class and all other students were heterogeneously grouped. In the control group, *all* students were heterogeneously grouped. Seventy-five matched pairs of rapid learners and 193 pairs of non-rapid learners were studied over a two-year period. At the end of the interval, the experimental group of rapid learners showed more progress in academic achievement in all areas measured by the *Iowa Tests of Basic Skills* than did the control group, but only reading achievement and composite scores were significantly different; the experimental group of non-rapid learners showed more growth than did the control group of non-rapid learners in all areas of academic achievement except vocabulary, but none of the differences were significant.

DeGrow (1963) conducted a study in Port Huron, Michigan, involving a three-part research design. The criterion was reading achievement as measured by the *California Achievement Tests.* In a one-year study, two groups of students in grades 4, 5, and 6, matched on the basis of IQ, grade level, sex, and reading scores, were involved. One group was taught in a homogeneous setting, with vertical grouping* according to reading level; the other, in heterogeneous classes. At the end of the year, there were no significant differences in achievement between the homogeneous and the heterogeneous groups, even though variation in reading grade equivalents had been reduced from 8.0 to an average of 1.13 through the homogeneous grouping. In a four-year cross-sectional comparison, comparative data collected for two preceding years indicated that vertical grouping did not make a difference in the average reading achievement gains of students. In a three-year longitudinal comparison, mean reading gains for 180 students who had remained in the homogeneous groups through grades 4, 5, and 6 were not related to this method of grouping. It was DeGrow's conclusion that vertical ability grouping in reading in grades 4 through 6 did not contribute to gains in reading achievement.

Kline (1963) evaluated the tracking plan in St. Louis public high schools. An experimental group was tracked over three to four years, while a control group was traced through their results in heterogeneous classes over the same period. The two groups were matched initially on the *Iowa Tests of Basic Skills.* The final criteria were teachers' marks and scores on the *Iowa Tests of Educational Development.* On teachers' marks, 40 experimental-control comparisons were made. For four of the 40 comparisons, the experimental group was higher; for five of the comparisons, the control group was higher. On the tests there were 36 experimental-control comparisons. For four of these, the experimental group was significantly higher; for seven, the control group was significantly higher. Kline concluded that tracking appeared not to make much difference in the achievement of St. Louis public high school students.

A group of sixth graders who had been in homogeneous (ability-grouped) classes for a three-year period were compared by Morgenstern (1963) with a group that had been instructed in heterogeneous classes over the same length of time. The measures used were the *Stanford Achievement Test,* the *California Test of Mental Maturity,* and two tests of personal and social adjustment. While Morgenstern's major conclusion was that ability grouping does not result in significantly greater increments in overall academic achievement than does heterogeneous grouping, one of her important subfindings was that in certain specific subject areas, such as language and word meaning, the homogeneous group was significantly superior; another was that for the lowest IQ groups, those grouped homogeneously showed greater gain in academic achievement. Her findings regarding personal and social development are reported later in this section.

Tobin (1965) reported a study involving students from grades 1 through 6. The study, covering an eight-year period, included a heterogeneous control year, 1954; a transition year, 1955; and six experimental years, 1956-1961. During the experimental years, students were grouped yearly within each grade on the basis of reading ability; similarly, each year the

*For vertical grouping, students in grades 4, 5, and 6 were assigned to reading classes on the basis of reading level rather than by grade.

grades were divided into thirds on the basis of IQ. Each experimental year was compared with the control year. Tobin found that the total group, each of the three IQ level groups, and every separate grade maintained stability in mean intelligence over the eight years. The total group showed positive upward trends in reading and in general achievement; the same was true for the high, average, and low ability students. There was an upward trend in general achievement that was significant in all grades except the third. For reading, gains were significant in grades 1, 2, and 6. Tobin believed that there was no Hawthorne Effect in his study, inasmuch as the greatest increases took place in the later years of the study.

A number of single-year studies of ability grouping, some involving single grades and/or single subjects and others involving several grades and several aspects of student achievement, have been reported in the past ten years. Studies by Provus (1960). Fick (1962), Loomer (1962), Mikkelson (1962), Drews (1963), Flowers (1966), and Peterson (1966) have been selected for review here.

Provus (1960) studied 494 students in grades 4 through 6 in Homewood, Illinois. Homogeneous classes made up of students grouped for arithmetic only—the academically talented, average students, and slow learners —were compared with heterogeneous classes. On the basis of results on the arithmetic concepts subtest of the *Iowa Tests of Basic Skills,* Provus concluded that children at all ability levels, grouped by ability, were more familiar with arithmetic concepts and fundamentals than children who were not grouped according to ability. He further concluded that the academically talented students profited most from ability grouping; the average students profited slightly; and the slow students profited no more from homogeneous grouping than they did from heterogeneous grouping.

Grade 7 students in Olathe, Kansas, were studied by Fick (1962). He formed homogeneous and heterogeneous classes which were pretested and post-tested with the *Iowa Tests of Basic Skills* and three measures of attitudes, values, and anxiety. Fick found that his students in homogeneous groups averaged no differently on achievement tests than those in heterogeneous groups. The low ability students in heterogeneous classes were superior to those in homogeneous classes in reading comprehension and punctuation. High ability students in homogeneous classes scored higher on uses of references than did those of similar abilities who were taught in heterogeneous classes. The homogeneous-heterogeneous comparisons on the other instruments will be discussed later in this section.

Loomer (1962) conducted a study involving 490 students in grades 4, 5, and 6, enrolled in 23 different classes. Five heterogeneous classes contained all levels of ability. The homogeneous groups included a high group and a low group. The homogeneous high group contained all ability levels except low students; the homogeneous low group contained all ability levels except bright pupils. The achievement growth from February of one year to February of the next year was measured by the *Iowa Tests of Basic Skills.* Loomer reported no significant differences between homogeneous high and heterogeneous groups except for vocabulary at grade 5, in which the homogeneous high group was superior. No significant differences were found between homogeneous low and heterogeneous groups. No significant differences between homogeneous high and homogeneous low groups were found except in grade 4, in language and total achievement, and in grade 5, in vocabulary and total achievement, where the homogeneous high arrangement produced superior results. No significant differences were found on any test between homogeneous and heterogeneous classes insofar as bright level students were concerned; for the low ability students, the only significant differences in achievement were found in grade 5 in reading and in grade 6 in language, where the heterogeneous grouping proved superior. Loomer concluded that his evidence indicated no decided advantage to homogeneous grouping over a random method of assigning students to classes.

Mikkelson (1962) studied 280 students of superior mathematical ability in grades 7 and 8 in a Minneapolis junior high school. One hundred forty of the students were studied during the 1958-59 academic year; the other 140, during 1959-60. Thirty-five students in each grade, assigned to one homogeneous class on the basis of mathematics achievement, Otis IQ, and teacher judgment, comprised the experimental group; the control group was comprised of 35 students placed in traditional heterogeneous classes in each grade. During the first year, no special adjustment in curriculum was made; in the second year, the curriculum was adapted to the homogeneous group by means of acceleration. Mikkelson reported that no differences in mathematics achievement resulted from grouping students of superior mathematical ability when no adjustments were made in the teaching procedures or the curriculum; but that with an accelerated curriculum, the homogeneous group accomplished more than those regularly grouped.

In a one-year study of student abilities, learning patterns, and classroom interaction, involving 432 ninth-grade English students in four schools in Lansing, Michigan, Drews (1963) worked with academically talented, average, and slow learners assigned to homogeneous and heterogeneous classes on the basis of IQ and reading and language skills. Teachers were

matched so that each grouping level had an equal number of experienced and inexperienced instructors. Tests administered at the beginning and end of the school year revealed no significant differences in reading and language achievement, problem solving, and critical thinking between homogeneously and heterogeneously placed students at any ability level during the year.

Flowers (1966) tested what is commonly called the "self-fulfilling prophecy."* He hypothesized, "If one of two groups of students of similar tested ability and achievement is assigned arbitrarily to a moderately higher level section and is taught on that level for a year, the group so placed will surpass the other group in tested achievement by the end of the academic year." Flowers worked with seventh-grade students in two experimental groups and two control groups matched on scores on achievement and intelligence tests. The two experimental groups were shifted to higher section designations than their test data would have warranted without their knowledge or their teachers'. Despite a slight trend to higher achievement for the experimental groups, Flowers concluded that his hypothesis was not validated. Extraneous uncontrollable factors evidently operated in this research, such as community differences, school assignments, and teacher styles. It appeared possible to Flowers that the upward trend was related to teacher expectation since a questionnaire indicated that teachers of the experimental groups favored the "high" ability groups, were more sensitive to the need for remedial instruction, and made greater attempts to motivate the "high" ability groups.

Peterson (1966) studied students in grades 7 and 8 in a junior high school in Chisholm, Minnesota. These students were grouped in three ability levels—high, middle, and low—on the basis of six tests of scholastic ability. One half of the students at each level were taught in homogeneous groups; the other half were placed in matched heterogeneous sections. Eight achievement tests were given at the beginning and the end of the year in order to measure growth. At the end of the year, Peterson studied differences in the groups in achievement and attitudes toward school. All comparisons that showed significant differences between the groups—and the majority of these were for arithmetic achievement—favored the heterogeneous groups; but only three of the 24 comparisons at grade 7, and eight of the 27 comparisons in grade 8, were statistically significant. Peterson concluded that his study "failed to offer sufficient support for the superiority of either homogeneous or heterogeneous grouping."

It is interesting to note that while the great debate has been going on in the United States during the 1960's over the relative merits and demerits of ability grouping or "tracking," a similar debate has been taking place in England over their ability grouping or "streaming" system. Since, however, most of the significant research that has been done in England has been concerned with the effects of "streaming" on the social and personal development of children rather than on their academic attainments, the pro's and con's of streaming, as the English see them, will be discussed later.

* * * * *

A brief summary note regarding the effects of ability grouping on school achievement is that (1) separation into ability groups, when all children involved are considered, has no clear-cut positive or negative effect on average scholastic achievement, and (2) the slight trend toward improving the average achievement of high level groups is offset by a substantial loss by average and low groups. How these effects may be produced by the fact of ethnic and socioeconomic separation resulting from ability grouping is the subject of a later part of this section.

One special footnote is a trend in the results of ability grouping nowadays as contrasted with findings in the 1920's and 1930's. The earlier studies more often than not reported gains by the low groups and losses by the high groups when compared with similar students taught in heterogeneous classes. Today, the trends are just the opposite: any advantages are shown by high level groups; disadvantages are shown quite commonly for the low groups. Why?

A possible explanation is that in the earlier period strong academic motivation was accepted as a favorable characteristic of individuals, to be prized when noted, but not to be expected under the prevailing drill emphasis in instruction, while the current concept of a "dropout" as one deprived unfairly was yet to be born; currently, since Sputnik in 1957, strong academic motivation and achievement have been "demanded" by our technological society, especially through middle-class parents, with concomitant wide acceptance that lack of this composite of achievement and motivation in minority groups is a fundamental source of deprivation. The "low" feel low and behave

*Heathers (1969) cites the study by Rosenthal and Jacobson (1968) as the most dramatic evidence of the self-fulfilling prophecy. In that study, randomly selected students from a class were identified to the teacher as "academic spurters." Over the next several months, these students showed reliable gains in IQ scores, a finding that was equally true of students who were in fast, medium, or slow groups. Unfortunately for this viewpoint, that study and further argument by Rosenthal (1969) involve questionable statistics (Thorndike 1968, 1969) and several efforts at replication have proved unsuccessful (Barber et al., 1969).

ineffectively to secure the benefits in upward mobility that education provides.* All of which leads naturally to the discussion of the impact of ability grouping on and through affective development.

GROUPING PRACTICES AND AFFECTIVE DEVELOPMENT

Many opinions have been hazarded concerning emotional and social results of ability grouping, but the research evidence, at least until very recently, has been thin indeed, perhaps because emotional and social growth are more difficult to assess than intellectual growth.

Of the 33 studies reviewed by Ekstrom (1959), only one touched upon the social and personal adjustment of homogeneously grouped students. Byers (1961), reviewing the literature from 1930 to 1960, found only eight studies having to do with emotional and social growth, made prior to 1960, that were worthy of review. Borg (1966) included among his references eight studies made prior to 1960 that were concerned with non-cognitive variables; most of these were the same studies reported earlier by Byers. Of the 50 abstracts of research studies made since 1960, presented in the *NEA* Research Summary 1968-S3, 15 are concerned, in whole or in part, with social and personal adjustment. The contributors to the *Encyclopedia of Educational Research*—Otto (1941, 1950), Goodiad (1960), and Heathers (1969)—have had little to report on the relationship between grouping practices and affective development. Even Heathers lists fewer than a half dozen research studies concerned with this aspect of grouping.

As there has been little uniformity of opinion regarding the effect of ability grouping on the social development of students, just so has there been little uniformity among the findings reported for the research studies that have been made. However, while the literature concerning the social aspects of ability grouping includes at least some evidence to support any stand one might take, much of the evidence, especially the more recent evidence, seems not to support the generalization that grouping students according to ability contributes to the development of desirable attitudes and healthy self-concepts, especially among slow learners.

A number of the most significant research studies concerned with grouping practices and various non-cognitive variables—self-concept, attitudes, interests, sociometric patterns, personality traits—deserve to be noted here. Some of these have been reported by previous reviewers of the literature; a larger number have not. Because several of the studies were concerned with more than one variable, the studies are reported in chronological order rather than by aspect of affective development.

Research on Affective Impacts Prior to 1960

Luchins and Luchins (1948) interviewed 190 children in grades 4, 5, and 6 of a New York City public elementary school. They found that a high percentage of the students in the bright, average, and dull classes preferred to be, and believed their parents would prefer them to be, in the higher section of their grade rather than in the lower section. While most of those who were in the bright classes indicated that they would be unwilling to give up their higher class status even if the teacher of the lower class were "better and kinder," a majority of those in the dull and average classes would have been willing to change their class because of the teacher factor. A high percentage of the children in the bright group did not frequently play with, nor would they choose their best friend from among, students in the less able class; while most of those in the average and dull groups were willing to choose playmates from the brighter group and showed a willingness to select best friends without regard to the identification of their class. Many dull students felt inferior and ostracized, and believed that there was stigma attached to the dull class level. There was strong social pressure to be in the higher class. The brighter children, in turn, were, on the whole, snobbish in their attitude toward those who were in the lower class. The Luchins concluded that homogeneous grouping seemed to help create a kind of caste system in the school.

Justman (1953) compared two groups of gifted high school students in New York City, matched on the basis of school attended, grade, sex, mental age, IQ, and achievement in reading and computational skills. The experimental groups were special rapid progress classes; the control groups were in heterogeneous normal progress classes. On the basis of results on a variety of tests, Justman concluded that segregation of gifted children in special progress classes is accompanied by academic achievement superior to that attained by matched students in normal progress classes with no detriment to social acceptance, interests, attitudes, and aspects of personality.

Horace Mann (1957) studied gifted children in grades 4, 5, and 6 in Pittsburgh. These children spent half of the school day with typical children in art, music, and physical education classes; the other half of the day

*Today, when "all the children of all the people" are in school up to a compulsory attendance age limit, the low achieving groups contain far more children of minority and low socioeconomic groups than earlier, when the comparisons were between groups within a narrower range of socioeconomic and ethnic variation.

was spent with other gifted children in classes devoted to academic learning and enrichment programs. Mann sought to determine how real were the friendships between gifted and typical children in this program of partial segregation; he also attempted to measure the social position of gifted children among their gifted classmates. He found that the gifted children chose as friends other gifted children more often than they chose typical children; typical children preferred other typical children as their friends. Rejections followed the same pattern. Mann concluded that grouping heterogeneously for part of the day did not produce the desired mingling among children of various ability levels. Acceptance and rejection were stronger within an ability group than between groups.

Luttrell (1958) studied 27 sixth-grade students in Greensboro, North Carolina, with IQ's of 130 or above in a special class (experimental), and a comparable group scattered among eight classrooms (control). Both the experimental and control groups were tested in the fall and the spring with an achievement test, the *Mental Health Analysis Scale,* and the *Social Traits Rating Scale.* The results on the *Mental Health Analysis Scale* showed no difference between the two groups, both groups making a slight gain during the year. On the part of the *Social Traits Rating Scale* based on teacher ratings, the groups were highly similar in November, but by May the control group showed greater incidence of these undesirable traits: boastful, bossy, noisy, sulky, quarrelsome. The part of the scale filled out by the students revealed a high degree of acceptance of the gifted child in the regular classroom. While the number of students was small and the time involved in the study short, the results generally favored the homogeneous group.

Goldworth (1959) studied a program in which gifted children in grades 4 through 8 in a suburban community in the San Francisco Bay area were assigned to special grouping for three hours a week. The 63 classrooms containing fast learners were randomly divided by school and by grade level into experimental and control groups which were comparable in size, IQ distribution, number of learners, and "degree of acceptance." Pretests and posttests, including the *Columbia Classroom Social Distance Scale* and three sociometric tests, were administered to all students. Goldworth found that the program had a limiting effect on the number of classmates whom children accepted as best friends, but had no effect on fast learners' acceptance of classmates as best friends, on group cohesion, or on subgroup preferences. The proportion of children who showed an increase in the degree to which they were accepted as friends by their classmates was significantly greater in the control groups. While this study is widely referred to in the literature, the results should be interpreted with caution since they were based on a study of somewhat less than five months in duration.

Research on Affective Impacts from 1960 to the Present

"Is ability grouping good in the way children look at themselves?" "Is it good in the way teachers look at children?" Maxine Mann (1960) studied 102 fifth-grade children through the use of self-reports. The children had been classified into four ability groups upon entering first grade on the basis of results on group intelligence tests and reading readiness tests, but were officially labeled only by teachers' names. Two of the questions children were asked to answer were pertinent to the study: "Which fifth grade are you in?" "How do you happen to be in this particular fifth-grade group rather than some other?" Mann found that the highest and lowest groups were most aware of the level of grouping, identifying their groups as "high fifth," "high," "best," "top fifth," and as "low fifth," "low," "lower," rather than by the teacher's name. The reasons the children gave for their assignment to their particular groups helped to bring their self-pictures into clearer focus. "I'm smart," "We're smarter," "I'm too dumb," and "We dont' know very much," "We are lazy" account for more than half the answers to the second question. In the top section, all the children gave positive responses in terms of ability or achievement and no negative responses. In the second section, all the responses were still positive although only about one fourth of them were in terms of ability or achievement. Most of the children in the third section and all of the children in the lowest section gave responses that indicated negative or unfavorable self-concepts. Mann's deduction was that ability grouping is cruel to all but the top students.

In a study of gifted children in California, Simpson and Martinson (1961) administered the *California Psychological Inventory* to 115 students in special class groups and 56 comparable students given classroom enrichment or acceleration at the eighth-grade and high school levels. The special classes made significant gains in 19 instances and significant losses in three instances on the Inventory, while the other students made significant gains in nine instances and significant losses in eight. Eighth-grade boys in the special classes made significantly greater gains than the other boys in Self-Acceptance; eighth-grade girls in the special classes made significantly greater gains than the other girls in Self-Acceptance and Flexibility; high-school boys in the special classes made significantly greater gains than did the other boys on Social Presence and Tolerance; and high-school girls in the other groups made significantly greater gains than the special class groups in Social Presence.

Fick (1962), in his study of seventh-grade students in Olathe, Kansas, previously cited (p. 29), used the *Index of Adjustment and Values,* the *General Anxiety Scale for Children* and the *Test Anxiety Scale for Children,* and the *Scale of Attitudes toward the School Situation*, along with an achievement battery. Classes grouped homogeneously and heterogeneously were pretested and posttested with all four instruments. As with achievement, the homogeneous and heterogeneous comparisons showed no significant difference in changes in peer behavior, learning needs, teacher-pupil relationships, or self-concept. Responses of students to the anxiety scales, however, indicated significant increases in both general and test anxiety on the part of the ability-grouped students.

In a study described earlier in this section (p.29), Drews (1963) used two self-concept measures. One instrument was the *Ability Self-Concept Rating,* consisting of a single question asking the student to compare his ability with the abilities of his classmates and to rate himself as above average, about average, or below average; the other was the *Concept of Self-As-A-Learner Scale.* a 20-item instrument developed by Drews from Bills' *Index of Adjustment and Values.* The *Ability Self-Concept Rating* was administered both as a pretest and as a posttest; the *Concept of Self-As-A-Learner Scale* was administered at the end of the study only. On the *Ability Self-Concept Rating* administered as a pretest, the one significant difference favored slow students in the homogeneous group; on the same instrument administered as a posttest, superior students in the heterogeneous groups and slow students in the homogeneous groups made significantly higher scores on the instrument. On the *Concept of Self-As-A-Learner Scale,* Drews found that although heterogeneously grouped superior students obtained higher mean scores, the differences were not significant.

Morgenstern (1963), it may be recalled (p.28), compared sixth graders who had been in homogeneous classes for a three-year period with a group that had been in heterogeneous groups over the same length of time. In addition to an achievement test and the *California Test of Mental Maturity,* she administered the *California Test of Personality* and *Thinking About Yourself.* As with achievement, ability grouping did not seem to result in a significantly better personal-social adjustment than did heterogeneous grouping. For students of average IQ. the better personal-social adjustment was found for those grouped heterogeneously.

In a study of homogeneously and heterogeneously grouped students of below-average ability in grades 7 and 8 of two Minnesota junior high schools, Torgelson (1963) administered the *Mooney Problem Check List* in addition to measures of achievement. On the Check List there was only one significant difference—from beginning to end of year the homogeneous group had a greater decrease than did the heterogeneous group in problems concerned with Home and Family. There were no significant differences between the two groups on sociogram results or in satisfaction with the classroom situation. Torgelson concluded that homogeneous grouping for below-average high school students was not superior to heterogeneous grouping.

Wilcox (1963) studied 1,157 eighth-grade students in 16 schools in five central New York State counties to determine the multiple effects of grouping upon the growth and behavior of junior high school students. The schools were selected to reflect wide variations in grouping practice; the independent variable used was degree of homogeneity of grouping by mental age in the several schools. In addition to instruments designed to measure mental ability. level of achievement. and critical thinking ability, Wilcox used the Maslow *Security-Insecurity Inventory,* a specially developed *Inventory of Attitudes toward Junior High School,* and an adaptation of the *Ohio Social Acceptance Scale.* He found that, for the total group, self-concept was unrelated to grouping; but for groups in the category below 90 IQ, there was a more positive self-concept with homogeneous grouping. There were no significant differences in attitude toward school when the total population was examined; but for students with IQ's below 105, attitude toward school was more positive under homogeneous grouping, and for students of high socioeconomic status who had IQ's of 105 or higher, it was poorer under homogeneous than under heterogeneous grouping. Wilcox concluded that, in the absence of curricular differentiation, homogeneous grouping has a significant positive effect upon the attitudes of low normal and low ability students toward self, school, and peers and a significant negative effect upon the attitudes toward self, school, and peers of high ability students from upper socioeconomic homes.

Adkison (1964) studied attitudes about self and group through the use of a questionnaire he developed, and administered in October and again in May to students in grades 3 through 6 in four schools, two at upper-lower and two at upper-middle socioeconomic levels. At each socioeconomic level, the usual heterogeneous grouping was used in one school, and homogeneous high and low ability groups, based upon test scores and teachers' judgment, were used in the other. His findings indicated that low ability students manifested less positive attitudes than high ability groups, the difference being greater with homogeneous groups than with heterogeneous classes, and greater at the upper-middle socioeconomic level than at the upper-lower socioeconomic level. Teachers in homo-

geneously grouped schools tended to favor such grouping, 44 percent to 31 percent; all who opposed homogeneous grouping were teachers of low ability classes. Adkison concluded that "Homogeneous grouping... appeared to be detrimental to those in low status groups and to have a positive effect on those in high status groups.... The evidence supports the concept that decisions to separate children through formal grouping patterns should include the question of values."

Bacher (1964) studied 60 slow learners in grades 6 through 8 in a New Jersey suburban school system. Thirty of the students were in two special classes, which served as the experimental group; 30 were in regular classes, which served as the control group. The *Columbia Classroom Social Distance Scale* and the *Davidson—Lang Check List of 35 Trait Names* were given at the end of the year, and a standardized reading test was given at both the beginning and the end of the year. Bacher found no experimental-control differences in self-concept or reading growth. However, social adjustment of the special-class slow learners was significantly more positive than that of the slow learners in regular classes. From this study, Bacher inferred that there is greater acceptance of peers by peers among slow learners in a special class than among slow learners in a regular class.

Deitrich (1964) made a comparison of the sociometric patterns of sixth-grade students in two school systems, one of which used ability grouping and the other, heterogeneous grouping. He found that no appreciable differences existed in the selection of friends between ability-grouped and heterogeneously grouped classes, that is, that ability grouping did not necessarily limit a child in his friend relationships. A strong tendency toward the "bright"selecting the "bright" and the "dull" selecting the "dull" as friends was noted; this was especially true when mutual friendships were involved. He also found that students do not necessarily choose bright students for help with difficult lessons, nor do they always choose a close friend for such help. Deitrich's study indicates that there are no appreciable differences discernible in the sociometric patterns of sixth-grade students who are grouped either heterogeneously or homogeneously.

Dyson (1965) studied two seventh-grade populations similar with respect to age, intelligence, academic achievement, school grades earned, the school environment which they experienced, and the socioeconomic levels of the communities in which they lived. The populations differed in the manner in which they were grouped for instruction. One group was instructed in a school in which students were assigned to classes heterogeneously; the other group, in a school which made a definite attempt to place learners in class sections that were homogeneous with regard to academic learning ability, IQ scores, achievement test scores, evaluations by sixth-grade teachers in the areas of reading and arithmetic, and the principal's evaluation of standing in class. The heterogeneously grouped students numbered 323; the homogeneously grouped, 244. Each of the groups responded to two instruments: the *Index of Adjustment and Values,* which yields an index of acceptance of self, and the *Word Rating List,* designed to yield an index of the more specific academic self-concepts. Dyson found that neither the patterns obtained when acceptance-of-self reports were compared with how students were grouped nor those obtained when academic self-concept reports were compared with how students were grouped varied from those to be expected as a result of random variation. He also found that while high achievers did not report significantly different patterns of acceptance of self from those of low achievers either in homogeneous or heterogeneous groupings, they reported significantly different patterns of academic self-concept from low achievers in both heterogeneous and homogeneous grouping situations. Dyson concluded that ability grouping alone did not appear to have a significant effect on either reports of acceptance of self or academic self-concept.

Zweibelson et al. (1965) studied the attitudes and motivation of approximately 360 eighth- and ninth-grade students assigned to three ability "tracks." An attitude survey with seven scores and a motivation inventory were administered before and after exposure to a program of team teaching. Contrary to expectations, the pretesting showed the brighter students in high ability groups tending to have significantly lower motivation scores than students in the lower ability groups. Students in the high ability groups also tended to have more negative attitudes toward group and school. There was little change in these basic relationships after exposure to the team teaching program; there was, however, at this point a significant positive relationship between the total attitude score and the motivation score not present originally. Zweibelson suggested that ability grouping may create more tension or pressure for the more able student, and that negative attitudes and lower motivation are possible consequences of this.

In the longitudinal study described earlier (pp. 26 ff.), Borg (1966) examined a number of non-cognitive variables at various grade levels in addition to achievement: sociometric choices, student attitudes, student problems, self-concept, and personality. During the four years covered by the study, he administered many different non-cognitive measures to different groups at different times. In reporting his study, Borg indicated that the net effect of ability grouping on af-

fective development was probably harmful to at least some of the students educated under such a system; and that where ability grouping showed any advantage over random grouping, the advantage was usually a slight one. In ability-grouped classrooms at the elementary school level, superior students showed a significant loss in sociometric status while average and slow students made gains in status. At the junior high level, ability grouping was consistently related to fewer problems. Attitude toward peers was found to be consistently related to ability in the randomly grouped classrooms while no such relationship was found in the ability-grouped classes. At all levels and for all samples, ability grouping was generally associated with less favorable self-concept scores. With respect to level of aspiration, Borg found no significant differences for students at the same ability levels in his randomly grouped and heterogeneously grouped samples; neither did he find that ability grouping led to a greater feeling of belonging on the part of students at any ability level, but that, instead, it provided a less favorable climate. His personality measures showed that the two grouping treatments did not affect differentially such personality variables as poise, ascendancy, and self-assurance, except in the case of students of average ability, where the random group showed a tendency toward more favorable scores. The Borg data suggest that the method of grouping students is not a uniformly significant factor in the feelings either of superiority or inferiority among elementary and junior high school students. The fact that self-concepts were lower for all groups at all levels, and that Borg himself questioned whether any small advantages to some compensated for the harmful effects on others, leads us to interpret his findings in this area as essentially negative.

Borg and Pepich (1966) conducted a controlled study of slow-learning tenth graders (IQ between 70 and 90) in a Salt Lake City high school. Students were matched for social class factors and grouped in English classes. Two different classes were studied in two different years; tests were administered at the beginning and end of each school year. The homogeneous grouping resulted in more class participation and more quality contributions. No significant differences were found between groups in either self-concept or attitudes; the only difference between groups was that the number of unexplained absences was significantly higher in homogeneously grouped classes. The authors concluded that the advantages of the more comfortable competition provided in homogeneous groups were outweighed by the disadvantages of the low-group label.

As part of their comprehensive study of the effects of ability grouping, Goldberg et al. (1966) reported student appraisals of their present status and their ideal or wished-for status on a variety of personal characteristics and abilities, as well as on academic expectations and satisfactions. Among the instruments used were *I Guess My Score* and three measures based on the method and format of the *Index of Adjustment and Values.* Although the presence of both gifted and slow students had statistically significant effects on the self-attitudes of the other ability levels, the results were inconsistent. The presence of gifted children tended to result in improved self-attitudes for brighter students and in less positive self-appraisals for slower students, but had little effect on average students. The effects of the presence of slow students varied from one area of assessment to another and also from one ability level to another; the presence of such students was associated with higher expectations of academic success held by the very bright and average students, but there was lower success expectation on the part of gifted students. Little support was found for the notion that narrow-range classes are associated with negative effects on self-concept, aspirations, attitude toward school, and other non-intellectual factors. In general, the effects of narrowing the range or separating the extreme levels was to raise the self-assessments of the slow students, lower the initially high self-ratings of the gifted, and leave students at the intermediate levels largely unaffected. The slow students also showed greater gains in their "ideal image" when the gifted were absent than when they were present. While grouping appeared to have no negative effects on the self-concepts and school attitudes of students in this study, it must be noted that largely because of the requirement that each participating school have at least four entering-fifth graders with IQ's of 130 or higher, the schools included in the sample were almost all located in predominantly middle-class sections of New York City and that their populations were, as a result, relatively homogeneous with regard to social class; furthermore, the low ability group was of low-average rather than low intelligence and included few students with IQ's below 90. Even for this select population the authors conclude cautiously: "Ability grouping is inherently neither good nor bad, it is neutral. Its value depends upon the way in which it is used. Where it is used without close examination of the specific learning needs of various pupils, and without the recognition that it must follow the demands of carefully planned variations in curriculum, grouping can be, at best, ineffective; at worst, harmful."

Olavarri (1967) studied the relative merits of heterogeneous and homogeneous grouping in terms of the students' self-concepts under these two arrangements. The *Concept of Self-As-A-Learner Scale* was

used to secure the responses of ninth- and eleventh-grade students concerning how they felt after two years of homogeneous or heterogeneous grouping. Olavarri found that lower ability groups consistently indicated better feelings of self-worth in the homogeneous setting than in the heterogeneous one, while the top ability group responses showed only a slight favoring of the grouped setting. Olavarri concluded that "Apparently the stigma of group labeling was readily offset by the classroom atmosphere and process." The percentage of "successful grades" was significantly higher in lower ability English classes than in the heterogeneous classes, while the reverse was true for the top groups.

Willcutt (1967) attempted to find a practical way of handling individual differences in the junior high school mathematics program. The entire seventh grade, 240 students, of a midwestern junior high school was involved. Fifty percent of the students were assigned to experimental classes—one review level (low), two standard (average), and one in depth (high)—and 50 percent to the control group. The instructional program was one whereby students were continuously regrouped during the year on the basis of their proficiency in each of the eight different mathematics topics studied. Of the 120 students in the experimental group, only seven remained in the "in depth" class throughout the year and only six in the review class. Pretests and posttests in arithmetic were administered, along with a questionnaire designed to test changes in attitudes. While there were no significant differences in arithmetic achievement between ability-grouped and heterogeneously grouped classes, the flexible ability grouping did result in significant attitudinal changes favoring the experimental group.

A study by Borg and Maxfield (1967) was concerned with the long-range sociometric development of a sample of students first studied at Grade 4 (Borg, 1966) and followed through Grade 11 in this later project. Sociometric choice measures were obtained on an initial sample of 1,031 fourth-grade students and subsequently on students available from this initial sample at grades 5, 6, 7 and 11. Subsamples of about fifty students who had made the greatest gains and losses in sociometric status since Grade 7 were interviewed and administered an autobiographical questionnaire, a self-concept measure, a school attitude measure, and two personality inventories in Grade 11. Analysis of the data obtained indicated that the mean sociometric choice scores obtained at grades 7 and 11 by students in ability-grouped and randomly grouped classrooms were not significantly different at any of the ability levels. Differences in sociometric-choice patterns found at lower grade levels in the earlier study were not present at the secondary level. For four groups of students selected on the basis of scores obtained at grades 7 and 11 and identified as the Low-Low group, the High-High group, the Up group, and the Down group, none of the measures obtained in the earlier grades yielded differences sufficiently large or sufficiently consistent to be of any value in predicting future trends in sociometric status of elementary school students.

Sarthory (1968) studied sixth-grade students from six schools in a large metropolitan area in the Southwest. Three schools used heterogeneous grouping, and three used two homogeneous groups, one above and one below the school's median IQ. Varying proportions of Anglo- and Spanish-American students attended the schools. Self-concept was measured by the Sense of Personal Worth Scale of the *California Test of Personality;* intercultural attitudes were measured by a semantic differential test; occupational aspirations were measured by the Haller *Occupational Aspiration Scale;* and educational aspirations were assessed by the use of a five-point scale devised by Sarthory. The major findings were that "An ability group cannot be considered as a reference group. Rather, self-concept, intercultural attitudes, and aspirations appear to be based on one's membership in other social groups, particularly the family and socioeconomic status." According to Sarthory, grouping did not significantly affect these variables except for occupational aspirations: the grouped students of high IQ had higher aspirations than the ungrouped high IQ students. There were indications in this study that grouping tended to inflate or deflate slightly attitude sets which were grounded mainly in socioeconomic status and IQ considerations, and that intercultural attitudes were based more on socioeconomic status factors than on ethnic factors. Sarthory recommended that ability grouping not be used. He suggested, instead, the use of techniques of individual instruction, formal preschool programs to remove deficiencies, and the establishment of attendance districts to insure no "perpetuation of tensions of the larger society."

Good and Brophy (1969) reported observational data on treatment of boys and girls in first-grade reading instruction. They found that differential treatment by sex did not occur in the reading period, but at other times when boys' disruptive behavior drew more rebukes. These observational data were contrary to children's reports of teacher behavior; classmates did not make this distinction but, rather, indicated that teacher rebukes of boys were quite as excessive in reading periods as at other times. In a reworking of the same data, Brophy and Good (1969) found teachers gave more positive reinforcement to those children they judged most able and more negative or unresponsive reactions to those judged less able.

Ability Grouping—British Style

In order to serve the highly selective university system in England (only seven to eight percent of the young people of college age are at the universities), a sorting-out process has, over the past half century, resulted in rigidly "streamed," or ability-grouped, primary schools (ages 7+ to 11+), based on reports of infant schools (below age 7+) and internal and external examination, rigidly "streamed" junior schools (ages 11+ to 16), and separate grammar and secondary modern schools (terminal). Only since World War II have comprehensive schools at the secondary school level emerged. In the early 1950's articles criticizing streaming began to appear, and research on the subject began to be published in the late 1950's. In 1967 appeared the Plowden Committee Primary School Report, which recommended unstreaming in infant schools with the hope that it would spread to primary and junior schools. This hope has not as yet been substantially fulfilled; the latest figures show that 58 to 70 percent of the junior schools still practice some form of streaming.

Ogletree (1969) discussed the pro's and con's of streaming and reported on some of the more significant research. The arguments advanced by British school administrators and teachers are strikingly similar to those advanced for and against ability grouping in the United States. Ogletree reported that most of the research conducted in England indicated that students in lower streams possessed a sense of failure resulting in a consistent decline in morale, effort, and attainment. He offered the opinion that even if streaming gave sound and true homogeneous groups, it "ignores the more subtle aspects of the personality and the social aspects of man."

As indicated earlier in this section, few of the research studies concerned with the advantages and disadvantages of streaming have been concerned with academic achievement. Most have been concerned with the effects of streaming on the social adjustment and attitudes of students. Most of these studies suffer from the use of small samples and are, therefore, inconclusive; the best known studies that examine the effects of streaming on non-cognitive aspects show different results. With the research in Great Britain, as with the research in the United States, everyone can find evidence in previous research to support whichever side he takes on this issue.

Rudd (1958) tested the hypothesis that the attainments, attitudes, behavior, and personalities of students taught in a school organization based upon streaming would be influenced by that organization. His experiment involved two groups of 90 students entering the same school at the age of 11 years. The control group was organized into three heterogeneous classes whose membership did not change during the two-year period following entry to the school; the experimental group was organized into three streams and students were transferred between streams after each half-yearly examination. Neither tests of ability nor tests of attitude toward examinations, school lessons, and school life in general yielded significant differences between groups. Samples of classroom behavior revealed that in the group organized into streams, fewer social contributions were made by students and there was more aggressive behavior and less attention to work. Estimates of personality made by teachers revealed no significant differences between groups while students' self-estimates revealed an extensive, but probably temporary, deterioration in personality following regrouping. No general long-term effects attributable to streaming were discovered.

Cox (1962) investigated the effects that educational streaming practices have on scores on the *General Anxiety Scale for Children* and the *Test Anxiety Scale for Children.* He used an Australian adaptation of both scales, which he administered to a sample of 266 fourth- and fifth-grade children in two schools in Canberra. In each school, the children had been divided into "superior" and "inferior" subgrades on the basis of their academic records in the first three grades of school. Cox found that general anxiety scores were independent of educational practices but test anxiety scores were significantly, and negatively, related to level of subgrade. He also found that test anxiety scores increased with grade.

Willig (1963) investigated the social implications of streaming by academic attainment in the junior school with particular reference to its possible effects on (a) social interaction between children of differing intelligence and socioeconomic status; and (b) differences in social adjustment and social attitudes between children in streamed and unstreamed classes, and such differences between children in "A" (faster) and "B" (slower) streams. Two hundred boys and girls, aged between 9 and 10 years, were drawn from two contrasting social areas. In each area an "A" class, a "B" class, and an unstreamed class were studied. A sociometric test was administered to determine social interaction between the various criterion groups. The N.F.E.R. *Primary Verbal Test 1* was used as a measure of intelligence, and an index of socioeconomic status was provided by grading occupations of parents. Teacher ratings were obtained to determine incidence of maladjustment, and an attempt was made to measure children's social attitudes by means of a sentence completion test. Other measures included a brief questionnaire designed to explore children's attitudes toward streaming. Taken as a whole, the evidence from

the sample pointed to the social advantages of heterogeneous grouping as opposed to streaming by academic attainment. Heterogeneous grouping provided greater opportunities for the formation of mutual relationships between children of different intelligence and socioeconomic status levels. In streamed schools cleavage existing between "A" and "B" streams operated to force the more intelligent "B" class children of intermediate socioeconomic status to associate only with their intellectual and social peers, or with children in lower intelligence and social class groups. There was a tendency for children in unstreamed classes to be superior in social adjustment, as defined by Stott's *Six Adverse Adjustment Pointers* scale, a relatively crude instrument but one which successfully differentiated between the criterion groups at the 5 percent level of significance. It was also found that in streamed schools "A" class children tended to be superior in measured social adjustment and socioeconomic status to those in the "B" class. Since social interaction between streams was very limited, "B" class students were prevented from associating with the "better adjusted" "A" class children, who were more likely to conform to a generally accepted system of values. Finally, it was shown that children in streamed schools were fully aware of the advantages associated with "A" class status and of the inferior position of the "B" class in the school hierarchy.

Kellmer-Pringle and Cox (1963) studied 235 children who comprised the entire fourth year in two junior schools in the Midlands. One school was organized in a mainly adult-directed traditional form in which competition, streaming, and class teaching were emphasized. The other school maintained a child-centered progressive regime in which cooperation and the realization of each individual's potentiality was emphasized; in this school, neither streaming nor group tests of any kind were used until the last year in the school. The headmasters of both schools were convinced of the soundness of their approaches and both gave positive and strong support to the staff; each was reportedly dedicated to the welfare of the students. On both the *General Anxiety Scale for Children* and the *Test Anxiety Scale for Children*, children in the unstreamed, child-centered, progressive school received significantly higher mean scores (less anxiety) than those in the streamed, adult-directed, traditional school.

Levy, Gooch, and Kellmer-Pringle (1969) carried on a longitudinal study of the relationship between anxiety and streaming in two junior schools, one (School T) a traditional school with streaming throughout and one (School P) a "progressive" school with no streaming until the fourth grade. One hundred eighty-one boys and girls were involved. The *General Anxiety Scale for Children* and the *Test Anxiety Scale for Children* were administered on three equally spaced occasions over a 12-month period. The 11+ examinations* were taken between the second and third testing occasions. Although in some cases GA (general anxiety) and TA (test anxiety) scores yielded parallel findings, differences in school regime and interactions with this factor affected GA scores generally, whereas TA scores showed different relationships with streams on different testing occasions. In School P, GA was found in the lower streams, while in School T the lower stream had the highest mean (less anxiety); these results were broadly true for each testing occasion. The lower streams tended to show more TA, but this tendency differed in strength from one testing to the next. In School P, both scores fell on the second testing, but on the third occasion GA remained high whereas TA showed a fall. The investigators suspected that the onset of streaming and the coming of the 11+ examination aroused previously unexperienced anxieties in School P. The passing of the 11+ examination by the third testing might then be supposed to allow TA to fall, even in School P, while GA remained high in that school as a function of the continuing and widespread social effects of streaming.

Griffin (1969) studied 586 children at age 14+ in three grammar, three comprehensive, and six secondary modern schools. No systematic differences in educational attainment were found. Children in the comprehensive schools recorded better attitudes toward school; boys and girls in comprehensive schools, at each level of ability, expressed the wish to stay at school longer than did their counterparts in grammar and secondary modern schools although the differences were not significant at the 5 percent level. For children of average and below average ability, the comprehensive schools appeared to provide a more stimulating environment than did the secondary modern schools. If the grammar schools are considered to be upper level and secondary modern schools to be lower level, both homogeneously organized, and comprehensive schools to be heterogeneously organized, this study presents results that are similar to those being reported for a great many studies in the United States for homogeneous versus heterogeneous grouping.

Under the sponsorship of the National Foundation for Educational Research in England and Wales (N.F.E.R.), Bouri and Barker Lunn (1969) made a

*The 11+ examination was for a number of years administered universally in Great Britain at the end of the junior school to determine eligibility for secondary school education in the grammar school (academic) or the secondary modern school (terminal). While it is still widely used, it is not as popular as it once was. Critics maintain that it sorts too early and too permanently for many children.

study of the effects of different types of school organization on student achievement and behavior in 28 junior schools having four classes or fewer. The two main forms of organization were the Traditional Standard method, approaching the homogeneous, involving rough allocation of children to classes according to age but with double promotion of the more able students and retention of the less able, and the According-to-Age, or more heterogeneous, method, which adheres strictly to the criterion of age (in months) in the assignment of students. In schools with fewer than four classes, it is necessary to split a year-group of students and put more than one year-group in a class even in According-to-Age schools. Ninety-four teachers and 2,822 students were involved in the study. The two halves of the sample matched satisfactorily on nine out of ten criteria; suitable adjustments were made for the tenth criterion, father's occupation. Teacher ratings and sociometric data revealed no differences in total maladjustment ratings, although on individual traits certain differences emerged. For example, students from all social classes in Traditional Standard schools were considered by their teachers to be more prone to bullying and fighting, and students of the upper socioeconomic group in these schools were rated as more disobedient than their According-to-Age counterparts. On the other hand, students in lower socioeconomic groups in According-to-Age schools were considered more withdrawn and less pleasant to have in class. On the basis of sociometric data, classes in According-to-Age schools had a warmer and more friendly atmosphere.

The larger study conducted by Barker Lunn (1970) under the sponsorship of N.F.E.R. is easily the most extensive ever conducted to examine the effects of streaming and non-streaming on the personality and social and intellectual development of junior school students. A major part of the research was concerned with the follow-up, through their junior school course, of approximately 5,500 children in 72 junior schools, 36 streamed and 36 unstreamed. The students were tested initially at age 7, in 1964, and then annually until 1967, when they were in their final junior school year. The measurement instruments were tested and questionnaires designed to assess performance and attitudes in nine different areas: (1) attainment in reading, English, and mathematics; (2) verbal and non-verbal reasoning; (3) creativity, or divergent thinking; (4) interests; (5) school-related attitudes; (6) personality; (7) sociometric status; (8) participation in school activities; and (9) occupational aspirations. Information was also obtained on teachers' attitudes toward streaming and other educational matters on their classroom practices and teaching methods. In addition, a limited study was made of parents' attitudes.

One of the most important findings concerned the role of the teacher. Teachers within streamed schools were more united with respect to both their views on educational matters and their teaching methods; in non-streamed schools there was a wide divergence of opinion. About half the teachers in non-streamed schools held attitudes more typical of teachers in streamed schools; this group of teachers created a "streamed" atmosphere within their non-streamed classes, their teaching methods and attitudes tending to reflect the "knowledge-centered" pattern found in streamed schools rather than the "child-centered" pattern found in the non-streamed school. Because this could easily result in modifying the true effects of an educational policy of non-streaming, all analyses were carried out in terms of two teacher-types: Type 1 held attitudes and used teaching methods typical of non-streamed schools and Type 2 was typical of streamed schools.

The children's academic performance, in the main, was unaffected by their school's organization or their teacher's attitude toward streaming, although the attainment of children who were promoted or demoted was clearly affected, that of the one group favorably and that of the other group unfavorably. In general, neither school organization nor teacher-type had much effect on the social, emotional, or attitudinal development of children of above average ability, but they did affect strongly those of average and below average ability. Children of average ability were particularly influenced by teacher-type in the development of their teacher-student relationship and academic self-image. In these two areas, students who were taught by "typical streamers" in non-streamed schools held the poorest attitudes. Boys of below average ability also had the most favorable teacher-student relationship with typical non-streamed teachers in non-streamed schools; but more boys of below average ability also had a good academic self-image in streamed schools. In the development of certain school-related attitudes—attitude to class, "other image" of class, and motivation to do well in school—children of average and below average ability did better in non-streamed schools.

The number of streams in streamed schools appeared to be important. Although students in A-streams tended to improve and those in lower streams to deteriorate in their attitudes, the effect was more pronounced in the bottom streams of three- or four-stream schools.

Children in both streamed and non-streamed schools taught by teachers of either type tended to choose other children of similar ability and social class as friends, although there were a greater number of mixed friendships in non-streamed classes. There was little difference in social popularity of children between

those in streamed schools and those taught by "typical non-streamers" in non-streamed schools; however, more children of below average ability taught by "typical streamers" in non-streamed schools were friendless or neglected by other children. More children in non-streamed schools participated in school activities; but in both kinds of schools, especially the streamed schools, bright children and children from the higher social classes tended to be more active.

Although parents' educational aspirations for their children appeared to be influenced by the type of school attended, and in streamed schools by the stream-level, this was not true of the children's own occupational aspirations. Whether the desired occupation was based upon fantasy or otherwise, there was little difference between the choices of children in streamed and unstreamed schools. The aspirations of the boys seemed to be much more unrealistic than those of girls and ability had less effect on their choice.

* * * * *

Before attempting to summarize the evidence on the impact of ability grouping on the affective development of children on the present scene, a number of observations should be noted. First, studies of the impact of ability grouping on affective development are a more recent phenomenon than studies of impact on scholastic achievement. The studies in the 1920's and 1930's were concerned almost exclusively with the impact on achievement; the earliest study reviewed in the present section on impact on affective development is dated 1948. Second, many of the earlier studies —notably those by Drews (1963), Goldberg et al. (1966)—were concerned primarily with delineating the impact of ability grouping on "gifted" students in the period after Sputnik, when public concern was concentrated on cultivating high competence in mathematics and science, specifically stressed in the National Defense Education Act of 1958. The wording of conclusions of these studies points to concern with the affective development of the gifted when singled out for academic excellence and special opportunity; lower achieving groups are treated primarily as the norm group, the great remainder; comparisons are often with only the relatively low, around IQ 100. Third, as with studies of impact on achievement, the earlier studies show more benefits to the low achievers than now when the low achievers and the high achievers have ethnic and socioeconomic overtones.

On the current scene, then, the impact of ability grouping on the affective development of children is to build (inflate?) the egos of the high groups and reduce the self-esteem of average and low groups in the total school population. A new dimension of interpretation has been emphasized chiefly in the British studies of "streaming," where teacher attitude toward achievement is shown to have marked effect. In particular, teachers who bear attitudes of almost exclusive emphasis on academic achievement to the neglect of personal development exercise an especially pernicious influence on low-achieving children in heterogeneous classes where the differences are widest.

ABILITY GROUPING AND SEPARATION: ETHNIC AND SOCIOECONOMIC

Earlier in this section, it was shown that ability grouping has unfavorable effects on the scholastic achievement and the affective development of students placed in low groups, without redeeming benefits to match. To the extent that minority children are overrepresented in low ability groups, then, they are being made to suffer the unfavorable effects of ability grouping. Evidence is marshalled here which shows how sharply the minority children are separated from this stimulation by assignment to low, predominantly non-white classes in schools whose total student populations have been desegregated.

The Special Problem of Metropolitan Areas

First, it should be noted that the issue of desegregation and then resegregation by ability grouping is dead and meaningless in situations where inmigration of blacks and outmigration of whites to suburbs or private schools has already reached a point where the total local school population is predominantly black. The difficulties faced by a large metropolitan system's efforts to desegregate were examined in a study by Walker, Stinchcombe, and McDill (1967), who studied school desegregation in Baltimore.* These writers found that although both the Baltimore City system and the Baltimore County system have made some progress toward desegregation within each of the systems, when the two systems are considered as a single metropolitan system, no progress at all has been made. They point out that this is because, while segregation within the political boundaries has declined in importance, the county boundary has become the most crucial segregating influence in the metropolitan area; and unless integration can take place across the city-suburban boundary, neither school system, by itself, will be able to effect any appreciable amount of desegregation. They also point out the importance of private and parochial schools in maintain-

*In three journal articles variously authored by these three writers (1968, 1968, 1969), the separate points are outlined in briefer and more generally accessible form.

ing segregation. Even though concerted efforts might decrease segregation in the public schools, this would have relatively little effect because a very large part of the white school population who might go to school with blacks are not subject to public policy because they attend private and parochial schools.

The progress that has been made so far in the city of Baltimore has been made entirely by introducing blacks into previously segregated white schools; there has been virtually no introducing of whites into formerly all-black schools. Also, the fact that some schools which were previously desegregated have tended to become nearly all black is an indication that the number of predominantly black schools never declines; it always increases. The problem of resegregation has become a factor in the Baltimore schools. The only kind of desegregation that has apparently been implemented in Baltimore has been almost exactly equaled in recent years by a compensating number of schools which have become segregated. In the city of Baltimore, there are very few schools left which are still segregated white. These writers point out that, within a few years, it will be impossible for any city policy to achieve desegregation because there will be no more segregated whites to attend schools with blacks in an integrated environment. All of the above forces operate more strongly on the elementary level than on the secondary level; that is, more blacks go to school with whites in secondary schools than in elementary schools. Thus, desegregation progress has been more substantial and longer lasting in secondary schools.

In Baltimore, as elsewhere, the fundamental causative factor for segregation in the schools is the segregated pattern of housing within predominantly black or predominantly white neighborhoods. The elementary schools are almost exactly as segregated as are the neighborhoods in the metropolitan area of Baltimore. Senior high schools are considerably less segregated than the neighborhoods. This is an important aspect of the problem. Whatever influence the public school has on the level of segregation of social life in the city and county of Baltimore, it is more in the direction of desegregation than is true of neighborhoods.

One of the ideas examined in Baltimore was the notion of the "tipping point," that is, the proportion of blacks in a school beyond which whites will leave. The notion of the "tipping point" has been used in the city of Atlanta as an explanation of the tendency for schools which were all white at one time and then were desegregated to later become all black. According to the Baltimore study, the "tipping point" notion does not have validity in Baltimore. Instead of the "tipping point" idea, what is referred to is a demographic pressure in which an increasing black school population pushes about equally on all schools near enough to black neighborhoods for the children to go there. In the Baltimore situation, the fundamental aspect of neighborhood segregation is the differential net migration. As a black moves out of a desegregated neighborhood, he tends to be replaced by a black. The net migration, therefore, of whites into the metropolitan area takes place almost entirely in the suburbs, while the net migration of blacks takes place almost entirely by movement into the city. Differential net migration, therefore, constantly increases the blackness of inner city schools.

Viewed as a national problem, the problem posed by the Baltimore situation must be considered typical of virtually every large metropolitan area. The present situation there could be made to confer the benefits of desegregation on minority children only if the city and county schools were consolidated into a unitary school system and all private schools were also required to desegregate. What is said hereafter about ability grouping must be presumed to apply only to the situations outside metropolitan areas where predominant majorities are white, and blacks and other minority groups constitute absolute minorities when whole school districts are considered. In metropolitan areas, only drastic procedures of consolidating urban and suburban districts, and transportation of many students, would meet the requirement of equal access to educational stimulation for all groups.

Limited Research on Grouping Practices and Separation

As indicated earlier in this section (p. 26), relatively little attention to the consequences of ability grouping with respect to ethnic and socioeconomic separation is evident in the literature. There are a number of possible hypotheses to explain this omission.

One might argue, as has already been pointed out (p. 40), that the question as to the effects of a particular grouping practice on ethnic and socioeconomic separation is relevant only when the particular environment under study is ethnically and socioeconomically integrated; that is, given a community, school district, or school that is overwhelmingly segregated, it makes little sense to study the practical effect of grouping method X in relation to ethnic and socioeconomic differences in children—not that the question of *de facto* segregation is irrelevant or that it should not be of concern to educators and researchers, but that it is not a researchable question in a self-contained, racially isolated environment.

Further, given the degree of correlation between ethnic origins and socioeconomic class and performance on standardized measures of ability and achievement, to be discussed further later in this section,

it seems intuitively obvious, almost without the need for research, that a grouping practice that is based on such measures predetermines the placement of a high proportion of non-white and lower socioeconomic class children in the lowest homogeneous ability groups.

Finally, in the most recent examination of research studies addressed to the desegregated environment, Weinberg (1970) noted that in 1966 a Federal official in charge of desegregation enforcement replied to a Congressional inquiry as to the extent of research on desegregation: "The basic problem is there are few researchers that want to work on it for some reason...."

Notwithstanding the lack of scientific interest, it appears that the problem is probably more than a result of a fundamental dilemma in the American system: the isolation of certain ethnic and socioeconomic groups from the mainstream of a mixed society. Before, however, discussing other aspects of the problems and before presenting those few studies which document *de facto* separation in classrooms as a direct consequence of ability grouping, more extensive discussion of the extent of racial isolation is in order.

Racial Isolation in America

As reported by the U. S. National Advisory Commission on Civil Disorders (NACCD) (1968), there were 21.5 million Negroes in America in 1966. Fifty-five percent of this population lived in the South, 69 percent lived in metropolitan areas, and nearly half lived in 12 major cities. It is critical to note that, for Negores, inmigration to the cities has come to mean resegregation. According to *Racial and Social Class Isolation in the Schools* (RSCIS) (1969), prepared by the Division of Research of the New York State Education Department:

> Overall figures on urban centers do not reflect the segregation of Negroes within the cities. Like other immigrants, Negroes, as newcomers to the city, have lived in the oldest sections.... Once in the city, the Negro remains a city dweller. Economic limitations and residential restrictions have barred further movement. But, among the rest of the population, the trend for the past 25 years has been from the city to the suburbs. The combination of inmigration of Negroes and outmigration of white city residents has resulted in disproportionate numbers of Negroes in the cities in comparison with their representation in the total population. This disparity is intensified by the Negro birth rate and will become more pronounced. It is predicted that 13 major central cities of the country will be over 50 percent Negro in 1985.

With respect to the national school enrollment statistics, the inmigration of Negroes and outmigration of whites has had serious implications. For example, the NACCD reports that in the 1965-66 school year, 17 large city school systems in the nation (including seven of the ten largest) had Negro majorities in elementary schools. In only two of these cities, Newark, New Jersey, and Washington, D.C., did Negroes exceed 50 percent of the general population.

Even more serious is the finding that *within* a school system, Negro concentration in individual schools tends to be far greater than their proportion in the total enrollment. As reported in RSCIS:

> In 1965, in 75 major central cities, 75 percent of the Negro elementary pupils attended schools that were 90 percent or more Negro, while 83 percent of the white elementary children were in schools that were 91 percent or more white. These school systems were in both the North and the South, and the isolation of the Negroes held regardless of the proportion of Negroes in the total system.

These data tend to highlight a principal finding of the U. S. Commission on Civil Rights, reported in *Racial Isolation in the Public Schools* (1967):

> The causes of racial isolation in the schools are complex. It has its roots in racial discrimination that has been sanctioned and even encouraged by government at all levels. It is perpetuated by the effects of past segregation and racial isolation. It is reinforced by demographic, fiscal, and educational changes taking place in the Nation's metropolitan areas. And it has been compounded by the policies and practices of urban school systems.

As noted in the 1967 report of the U. S. Commission on Civil Rights, the policies and practices within the school system are seldom neutral in effect. Rather, they reduce, positively reinforce, or maintain ethnic and socioeconomic separation in the schools. Recent empirical studies clearly demonstrate how the educational policy of ability grouping tends to reinforce and, therefore, perpetuate ethnic and socioeconomic separation. In each of these studies, research is focused on a critical dimension of instruction: *the classroom composition of children.* Several of these studies are presented in detail later in this section.

Ethnic and Socioeconomic Status in Relation to Test Performance and School Achievement

Acknowledging that ability grouping as an educational policy is currently widespread and that student performance on standardized tests is frequently used as the criterion for classifying children into ability groups, then evidence bearing on the degree of relationship between ethnic and socioeconomic status

and achievement on standardized measures should be examined to determine the extent to which the practice of ability grouping is likely to separate children along ethnic and socioeconomic lines. The following summary does not claim to be an exhaustive presentation of the research bearing on the issue. Rather, it is intended to present some recent reviews of the literature which suggest that there is a clear relationship between ethnic and socioeconomic status and school achievement as measured by standardized tests, and to discuss the conclusions of a few of the most significant research studies.

If there is a paucity of research concerned with the relationship between ability grouping and ethnic and socioeconomic separation, there is no lack of studies concerned with ethnic origin and socioeconomic level in relation to performance on standardized tests. Numerous studies have been conducted on the relative performance of various ethnic and socioeconomic groups at the elementary, junior high, and senior high school levels. In all, the studies have used a wide variety of tests and measuring devices of school performance ranging from standardized ability and achievement tests, school grades, and teacher ratings, to highest school grade attained and average age for grade level.

Hubert Coleman, writing in 1940, was critical of studies done earlier. In his words:

> A review of earlier studies gives an inadequate and fragmentary picture of the relationship between socioeconomic status and such factors as intelligence, achievement, and personality adjustment. The studies show limitations such as small number of cases, lack of geographic sampling, questionable methods in the measurement of socioeconomic status and intelligence, incidental treatment of the socioeconomic factor, and homogeneous groups with respect to socioeconomic status.

Coleman himself (1940) studied data made available to him by the Advisory Committee of the Coordinated Studies in Education, Incorporated, on 4,784 junior high school students representing high, middle, and low socioeconomic levels as determined by a rating scale based on the Sims *Socio-Economic Score Card.* IQ's were determined by scores on the *Kuhlmann-Anderson Intelligence Tests* and level of achievement by scores on the *Unit Scales of Attainment* battery. Coleman found that differences in IQ favored the high socioeconomic group for boys and girls in each grade, with the median IQ falling between the two lower groups and tending to be closer to the lowest group. He also found a definite relationship between socioeconomic status and achievement favoring the high socioeconomic group. Coleman suggested that while his study showed a close relationship among socioeconomic status, achievement, and intelligence, it was not possible to say whether achievement is a result of socioeconomic status or intelligence, or to say that intelligence determines socioeconomic status or that socioeconomic status determines intelligence.

Dreger and Miller (1960) in a review of studies comparing Negroes and whites published between 1943 and 1958, stated that Negroes by and large scored lower on both traditional and so-called culture-fair tests of intellectual functions, but they noted that Negroes averaged well within the normal IQ range for whites.

Goldberg (1963) reviewed significant changes in recent decades that have created urgent problems for urban school systems. She also discussed the findings concerning achievement and motivation, with particular reference to Negro and Puerto Rican students. Claiming that, as a general rule, Negro children from low-income families achieved less well in schools than did comparable white children, she asked, "What accounts for the consistently lower academic status of children from disadvantaged ethnic groups, especially the Negroes, than of children from lower-class white families living in the Northern cities?"

Goldstein (1967), who presented an annotated bibliography of 80 studies made from 1938 to 1965, concerned with the education of urban youth of low income, wrote:

> It should come as no surprise to the informed reader that, by every conceivable measure, children of low-income families do not do so well in school as children from more affluent ones. The evidence has been presented in full and dramatic detail for the essentially white populations...; for the essentially Negro population...; for the mixed population...; and for cities in general.

Several sources suggest that social class status may have a greater influence on achievement than does intellectual ability as measured by standardized tests. McCandless (1967) summarized the data on the relative contributions of social status and intellectual ability to achievement and concluded:

> From the intelligence test differences between social classes, we would expect differences in school progress, middle- and upper-class children being expected to do better school work than lower-class children. The actual differences in academic achievement between social classes are even more dramatic than the differences in intellectual level. On the whole, lower-class children achieve less well in school than their intelligence tests predict they will, whereas middle- and upper-class children approach their academic potential more closely.

Most of the research studies of the relationship between ethnic and socioeconomic status and test

performance have resulted in findings similar to those already cited. Several additional studies of significance are summarized below.

Kennedy et al. (1963) studied 1,800 Negro elementary school children in the Southeastern United States to provide data on intelligence and achievement variables. The *Stanford Binet Intelligence Scale* was used to measure IQ, the *California Achievement Tests* to measure achievement, and demographic data not specified to measure socioeconomic level. The study resulted in the following conclusions: With respect to intelligence, the Negro children had a mean IQ of 80.7, but IQ was negatively correlated with age. IQ was highly correlated with socioeconomic levels though the differences were small between urban and rural residents. There was a significant difference in the mean levels of achievement test scores between the sample and the standardization group, and this difference increased with age. Achievement also correlated with socioeconomic level.

Deutsch and Brown (1964) explored intelligence test differences between 543 Negro and white first- and fifth-graders in different social classes, with particular focus on the lower class. The presence or absence of the father in the home was examined, and whether or not the child had had organized preschool experience. Social class was measured by a scale derived from prestige ratings of occupations as well as education of main breadwinners. IQ was measured by the *Lorge-Thorndike Intelligence Tests.* Differences between scores of Negro and white children were significant and were equally strong at all class levels. Negro children at each socioeconomic level scored lower than white children and Negro/white differences increased at each higher socioeconomic level.

With respect to secondary school, Goldstein (1967) noted a body of data, from Project Talent (Flanagan et al., 1964). Examination of these data in terms of socioeconomic differences tends to confirm the thesis that socioeconomic status is related to achievement. In this study, a two-day battery of tests and questionnaires was administered to 440,000 students in 1,353 high schools, "carefully selected to be representative of American secondary schools." The data indicated that, on the basis of a measure of general academic aptitude, males below the median were twice as likely as males in the top 30 percent to come from families possessing "only the necessities of life." Moreover, while over half of those in the lower 50 percent came from blue-collar families, less than one third of those in the top 10 percent did so. Rather, about 57 percent of the latter group came from white-collar families, while only 15 percent of the students in the lower 10 percent did.

In addition, Project Talent schools were classified into two relatively homogeneous middle- and low-income groups. One such group consisted of 27 schools that served predominantly middle-income students in New York City, Philadelphia, Detroit, Chicago, and Los Angeles. According to Goldstein, "there was virtually no overlap of the middle two thirds of the two populations, with low-income students consistently below middle-income students in the same school system."

Miner (1968) collected data from the files of 663 high school graduates in a midwestern city to investigate the relationships between a number of sociological factors, among them social class, family structure, and school achievement, at various periods in the child's academic career. Tests for which scores were available included the *California Test of Mental Maturity,* the *Iowa Tests of Basic Skills,* and the *California Achievement Tests.* Secondary school grades were also used. Significant relationships were found between a child's background and his early achievement. For the most part, the differences were small, but they were large enough to account for some of the variance in academic performance. Socioeconomic status was found to be positively related to the measures of performance.

In *Racial and Social Class Isolation in the Schools* (RSCIS) (1969), it was concluded that racial differences in achievement are approximately of the same order as the IQ differences between whites and Negroes. Data from the report *Equality of Educational Opportunity,* principally authored by Coleman (1966), based on a test of verbal aptitude, suggest an average difference in IQ of approximately one standard deviation between black and white children at grades 6, 9, and 12 in the Metropolitan Northeast. According to RSCIS, data from these grades also indicate a difference of approximately one standard deviation in the achievement levels of whites and Negroes of the Metropolitan Northeast. These deviation scores indicate that relative differences in achievement of Negroes and whites remain constant from grade to grade; grade equivalent scores indicate that these differences grow larger with successive grades. According to RSCIS, the interpretation of Negro-white achievement differences in grade equivalent scores as showing an increasing divergence with years in school is inappropriate for Negro-white comparisons. The conclusion reached in RSCIS was that the Coleman data, correctly interpreted (in standard deviation units), show that achievement differences between Negroes and whites do remain relatively constant from year to year.

Unfortunately for the purposes of this research, grade equivalent scores become progressively less meaningful in junior and senior high school; in fact,

the decelerating curve of growth on tests of basic skills might spuriously magnify differences expressed in grade scores. However, differences expressed in standard deviation units of white students of a given grade eliminate all opportunity to reflect increases in differences in average performance insofar as variability of individual achievement increases with age and schooling. The fact that grade score equivalents in the middle and upper elementary grades constitute approximately equal units and show progressively increasing differences between blacks and whites makes safest the interpretation that differences continue to increase, but in a fashion uncertainly represented by grade score equivalents.

Goldstein (1967) observed that although the instances have been few, some studies have come up with contrary findings. For example, Antonovsky and Lerner (1958) found that on the basis of a small class-matched sample of Negro and white students from lower socioeconomic status (complete data were available for 61 Negroes and 54 whites, about equally balanced for sex), the Negroes, despite greater handicaps, did as well academically as the whites, dropped out of school less frequently, and enrolled more often in the college preparatory program.

Goldberg (1963), in the reference previously cited (p. 43), cautioned:

> Despit consistent differences in demonstrated intellectual and academic ability... there is a great deal of over-lapping. In all studies there are some in the one group who resemble the other group far more than their own. And in all comparisons of lower- and middle-class children there is a sizable though smaller proportion of the former who score high on tests, do well in school, plan on advanced education, and have a high degree of similarity to the school performance of middle-class children. Conversely, there are middle-class children whose motivation and performance are poor indeed.

Despite some few exceptions, it appears from the above discussion that, for the majority of the population, ethnic and socioeconomic class variables consistently tend to be associated with school achievement as measured by widely used standardized tests. What does this mean with respect to the placement of children in elementary and secondary schools?

Empirical Consequences of Ability Grouping for Ethnic and Socioeconomic Separation in the Classroom

In view of the high degree of relationship between ethnic and socioeconomic status and performance both on standardized tests and in the classroom, it stands to reason that the use of ability grouping as a strategy for organizing children into classroom units should result in the separation of children along ethnic and socioeconomic lines. While, as has been indicated earlier, few research studies have been directed to separation along those lines, the studies that have been made show that such separation surely does exist, with children from the middle and upper classes found mainly in the middle and upper ability groups and children from the lower classes in the low ability groups.

In *Racial and Social Class Isolation in the Schools* (RSCIS) (1969), several studies are cited which show that grouping on the basis of achievement or aptitude tests leads to ethnic and socioeconomic isolation. Just as there are learning interference factors related to "inferior" schools, the report states, learning interference factors "should also be relevant in schools with grouping policies which result in either social class isolation within schools or combinations of different levels of racial and social class isolation, depending upon the class status of the white student population and proportion of 'integrated' Negroes in the school."

Heathers (1969), in his review of the literature on ability grouping, reported only four research studies concerned with the separation that can result from such grouping, none of them done in the United States. Despite the sparseness of research data, however, Heathers wrote:

> It is commonly recognized that low-ability groups in elementary school have a disproportionate number of boys, of children from lower class origins, and of children from minority groups. Ability grouping may thus be, in effect, an agency for maintaining and enhancing caste and class stratification in a society.

In the current search of the literature, several studies have been located which support the notion that ability grouping tends to isolate students of one ethnic group or socioeconomic level from another and that this isolation has deleterious effects upon various aspects of the development of students so separated. If, as a growing body of literature indicates, the impact of a school upon individual students is a function of peer interactions—that is to say, that students tend to learn as much from other students as they do from teachers—then these adverse effects can be anticipated.

Mehl (1965) studied 654 students in grades 5 through 8, who had been assigned to classes on the basis of group intelligence test performance from grade 4 on, to determine whether homogeneous grouping is an aspect of school procedure which may reflect, and thus reinforce, the social structure of the community. Social class was determined by Warner's *Index of*

Social Class scale. The same pattern of social class segregation was obvious in all four grades. Although all five social classes were proportionately represented in the two middle-ability groups, in the two top and two bottom groups there were statistically significant differences between the proportion of each social class level in the group and the proportion for the grade as a whole. Segregation was most pronounced in the extreme high and extreme low ability groups. A high relationship was found between measured IQ and achievement; a moderately low relationship was found between IQ and social class and between achievement and social class.

Wilson (1967) in a study of students in Richmond, California, found a marked relationship between the social class composition of schools and student performance. Regardless of their own social class, Richmond students were more likely to perform well in predominantly middle-class than in predominantly lower-class schools. When the relative importance of individual and school social class was assessed for black and white students separately, it was found that the student environment had a stronger relationship to the performance of black students than to that of white students. The performance of white students, although strongly related to the social class level of their fellow students, was more closely related to family background than was that of black students.

Wilson also weighed the effects of the social class composition of the school upon the same students over their entire elementary school careers. He found that in the primary grades the influence of the individual's social class was of great importance and that the social composition of the school was of little importance. However, over the period of eight years of school, the cumulative effect of the social class composition of the school increased sharply, so that in the eighth grade it was as significant as the individual's social class for student performance.

This pattern was generally the same where student attitudes were concerned, especially with regard to college aspirations and plans. College plans were found to be more frequent for both black and white students in schools with a higher social class level. Black students in schools of lower social class level, even though relatively advantaged, were less likely to attend college than similar students who were in school with a majority of more advantaged students.

In another "study" of the problem, Hobson vs. Hansen (1967), the basis question presented to the Court was whether the District of Columbia Board of Education unconstitutionally deprived the district's Negro and poor public school children of their right to equal educational opportunity with the white and more affluent school children. The case is directly related to the issue under discussion since it was the practical consequence of a track system which gave rise to litigation. Inasmuch as the court decision involves one of the most comprehensive discussions of every major issue introduced in this section, the relevant evidence presented to the Court will be presented in considerable detail.

The track system used in the Washington, D.C., schools was based completely on ability classification by standardized tests. Accordingly, students at both the elementary and secondary school levels were classified into separate, self-contained curricula or "tracks," ranging from "Basic" for the "slow" student to "Honors" for the gifted. The educational content ranged from the very basic to the very advanced according to track placement. In the elementary and junior high schools, three levels were used: Basic or Special Academic (for "retarded" children), General (for average or above-average students), and Honors (for the gifted). In the senior high schools, a fourth track (Regular) was added for college preparatory training of above-average students.

With regard to the pattern of socioeconomic separation occurring in the schools as a direct result of tracking, evidence submitted to the Court showed that when the high schools were grouped into three levels by median neighborhood income—high ($7,000 to $10,999), middle ($5,000 to $5,999), and low ($3,000 to $4,999)—the correspondence between track placement and income was exact. (See Table 9 below.) The economic-level correlations found in high schools were

Table 9

Percents of Students in Four Tracks in Washington, D.C., High Schools Serving Different Socioeconomic Levels of Neighborhood 1964, 1965

Median Neighborhood Income	Special	General	Regular	Honors
Over $7,000	0-7.4	7.8-43.7	46.1-80.0	10.2-17.1
$5,000-$7,000	4.7-9.9	39.0-57.7	32.9-49.2	3.2-7.8
Under $5,000	9.8-18.2	54.4-74.5	11.4-33.4	0-3.9

also found, generally, in junior high schools and elementary schools. The Court properly concluded that a student's chance of being selected for one of the higher ability tracks was "directly related" to his socioeconomic background.

With regard to the pattern of racial separation in the schools, the Court noted that, for a majority of District schools and school children, race and socioeconomic status were intertwined. The schools serving neighborhoods with income levels of $6,000 or below had Negro enrollments of well over 90 percent. The only predominantly white senior high school, serving a neighborhood of average income $10,374, had all but 8 percent of the students in Regular and Honors tracks in 1964 and 1965; no other school came close to that. A predominantly Negro school (90 percent) that was closest served a neighborhood with the third highest income level in the system ($7,650), but had 40 percent of its students in the lower non-college preparatory tracks. Of the six junior high schools having from 17 to 99 percent white enrollment in 1964, all six had Honors tracks; at least three of the schools were in the middle-income range. In six other middle-income schools, with student bodies better than 95 percent Negro, only three had Honors tracks on 1964, and this number dropped to two in 1965.

With reference to the distribution of track offerings in the elementary schools, only 16 percent of all Negro students were attending schools with Honors programs in 1965. Conversely, 70 percent of all white students had this advanced curriculum in their schools. This pattern of Honors track offerings in elementary schools also existed in the junior high schools.

Over and beyond the evidence presented above, the Court made a matter of record further data which illustrated how ability grouping practices result in the ethnic and socioeconomic separation of children. Looking at the racial breakdown of the enrollment in the Special Academic or Basic track, the Court noted that, at both the elementary and junior high school levels, the proportions of Negroes enrolled in the lowest track exceeded their proportionate representation in the total student body. On the other hand, the proportion of whites enrolled in the Special Academic track was significantly lower than the proportion of whites in the total school enrollment. It was clear that, as a general rule, in those schools with substantial numbers of both white and Negro students, a significantly higher proportion of Negroes than whites went into the Special Academic track (for "retarded students").

In summarizing the evidence, it was noted that the track system is by definition a "separative" educational policy, ostensibly according to students' ability level. However, the practical consequence of ability grouping is, by its application, to separate students largely according to their socioeconomic status and, to a lesser but observable degree, according to their ethnic status.

In recapitulating all the evidence and testimony, the Court pointed out the manner in which the concept and practice of ability grouping structures failure in black and lower socioeconomic class children, perpetuates unlawful *de facto* discrimination, and generally permeates an entire school system.

The point to be made here, it should be noted, is not to assess intent or blame. The finding is one of fact: that ability grouping produces segregation of students by socioeconomic status and, as a corollary effect, produces segregation by ethnic status. Insofar as such segregation has been shown to reduce stimulation of the low-achieving students to higher educational attainment, the effect of such ability grouping must be deemed to afford less than equal opportunity to the minority ethnic and lower socioeconomic groups.

Very dramatic evidence of how ability grouping based solely on test scores can effect decided ethnic and socioeconomic imbalance in the classroom is given by unpublished data made available by a Southern school district which was challenged in Court for its proposal to group black and white children in grades 3 through 8 in multiple sections on the basis of scores on tests in the *SRA Achievement Series*.

Recommended section assignments for children in Grade 5 in five subject matter areas are shown in Table 10 on page 48. Reading test scores for grades 3, 4, 6, 7, and 8, shown in Table 11, also on page 48, are typical of the scores in all five subject matter areas for these grades and, consequently, typical of recommended assignments.

After hearing testimony on the total plan for use of this ability grouping for organization of classes in the desegregated schools of the district, the Court ruled against the plan and in favor of a prior heterogeneous grouping plan with special instructional arrangements related to the disabilities being remediated.

Kariger (1962) studied the effect of an ability grouping plan used in the three junior high schools of a Midwestern city of 100,000 on socioeconomic stratification. In this plan, test scores were supplemented by teachers' and principals' judgment in making initial assignments to classroom groups and in making reassignments during the school year to correct for apparent misplacement by original assignment in the light of the subsequent academic performance of the students. Consideration of "teacher grades, study habits, citizenship and industry, social and emotional maturity" were allowed to guide these judgments. The tracking system called for placing those more than one grade advanced in the high track, those more than one

Table 10
Recommended Section Assignments Based on Battery Test Scores–Grade 5

	Reading		Mathematics		Language Arts		Social Studies		Science	
Section	Black	White	Black	White	Black	White	Black	White	Black	White
A	3	28	3	28	5	26	3	28	3	28
B	4	27	5	26	7	24	4	27	5	26
C	10	21	14	17	12	19	10	21	14	17
D	15	15	15	15	11	19	15	15	15	15
E	23	7	18	12	21	9	23	7	18	12
F	27	3	27	3	26	4	27	3	27	3
TOTAL	82	101	82	101	82	101	82	101	82	101

Table 11
Recommended Section Assignments Based on Reading Test Scores – Grades 3, 4, 6, 7, and 8

	Grade 3		Grade 4		Grade 6		Grade 7		Grade 8	
Section	Black	White	Black	White	Black	White	Black	White	Black	White
A	1	29	2	28	4	30	3	31	1	33
B	2	28	3	27	7	26	12	22	14	20
C	13	17	8	22	17	14	12	21	18	15
D	20	10	13	17	14	15	14	18	21	11
E	22	8	21	9	20	8	25	4	23	6
F	22	3	22	5	20	5	22	5	26	1
G	19	1	23	2						
TOTAL	99	96	92	110	82	98	38	101	103	86

grade retarded in the low track, and those less than one track above or below the norm in the middle group. Reassignments were often required to rectify class size, however.

In keeping with relations found quite uniformly in other studies, assignment to tracks on the basis of standardized test scores alone would have resulted in 77 percent of upper socioeconomic status children in the high track and only 38 percent of the lower socioeconomic status children in that track. Conversely, only 5 percent of the upper socioeconomic status children would have fallen in the low track while 26 percent of the lower socioeconomic status children would have been so classified. However—and this is the thrust of the study—80 percent of the upper socioeconomic status children whose test scores would have warranted placing them in the high track were actually in that track, while barely 50 percent (210 of 408) of the lower socioeconomic status children who qualified for high track placement on tests alone were so assigned. Children of the middle socioeconomic group fell into an intermediate position, 65 percent of those qualified by tests being assigned to the top track.

At the lower end, too few upper socioeconomic status children fell into the bottom track on test scores, so comparisons at that level can be made only between

middle and lower socioeconomic status children. Again, 37 percent of middle socioeconomic status children who qualified for the bottom track on test scores alone were placed in higher sections, while only 15 percent of lower status children whose test scores would place them in the bottom track were actually placed higher. To summarize, socioeconomic status of children significantly influenced track placement.

Turning now to the practice of reassigning upward children whose classroom performance reflected errors of too low placement initially, Kariger found that only 3.4 percent of students were affected; but, 70 percent of all reassignments were to higher classes. However, 93 percent of changes of upper socioeconomic status children were upward, 68 percent of middle group children reassigned were raised, and only 61 percent of the lower group changes were upward. The irony of it all is that the administrators were new to their schools and produced the initial separative socioeconomic effect without any history of prior bias of discrimination against the children based on experience with them.

A study of the Plainfield, New Jersey, school system was conducted by the Institute of Field Studies of Teachers College, Columbia University, to determine the practical consequences of the prevailing practices of ability grouping then in use at all grade levels. A 1967 statement of the Plainfield Board of Education expressed its policy in these terms:

> We recognize that within the Plainfield School System there are many different needs and opportunities for class and subject groupings. In order to meet these needs, there may be classes which can now be called racially imbalanced. It is our opinion that it is better to have such classes than not; that these classes should have an objective to prepare for the need for fewer such classes. We also recognize the opportunity for the display of ingenuity and innovation on the part of the staff to minimize any adverse aspects of such racially imbalanced groupings.

The effect of this policy is reflected in Hubbard Junior High School (1968-69), as shown in Tables 12 and 13. In Table 12, the data are for percents of the two separate ethnic groups in eighth grade to be found in the W (High) track, X (Middle) track, and Y (Low) track in each subject area. Table 13 gives the percents of total groups in each subject in each track in eighth grade that are black and white, respectively. All figures are to be compared to an overall total of 218 black and 90 white eighth graders, or 70.8 percent black and 29.2 percent white. Viewing the data either way, the whites are overrepresented in the top groups and the blacks are predominant in the bottom groups.

The upshot of this survey is significant. After pondering the evidence of ethnic segregation produced, the Board of Education took the following steps toward a more heterogeneous plan:

> To the extent possible, school principals in K-4 buildings have attempted to devise a planned heterogeneous grouping. In the spring, every teacher submits to the building principal a list of pupils in his class, noting whether each child 1) was reading at a high, average, or low level, 2) had been a discipline problem, 3) was Black or white, 4) was a boy or a girl. Using this information, principals attempt to develop self-contained classes composed of a "balanced" representation of children according to sex, race, and achievement, with discipline problems distributed as well.

Thus, the same test data used to produce homogeneous grouping can be used to define and establish heterogeneous groups. It remains to be seen how far and

Table 12

Percentages of the Hubbard Junior High School, Plainfield, New Jersey, Black and White Eighth-Grade Students, Enrolled in W, X, and Y Ability Groups by Subject Area, 1968-69

Subject	Race	Group W	Group X	Group Y	Total
English	Black	8.7	48.2	43.1	100.0
	White	58.9	34.4	6.7	100.0
Social Science	Black	10.6	46.8	42.7	100.1
	White	55.6	38.9	5.6	100.1
Mathematics	Black	3.7	56.9	39.4	100.0
	White	42.2	51.1	6.7	100.0
Science	Black	2.8	58.8	38.5	100.1
	White	43.3	50.0	6.7	100.0

Table 13

Percentage Composition of W, X, and Y Ability Groups by Race Hubbard Junior High School, Plainfield, New Jersey, Eighth-Grade Students (1968-69)

Subject	Group W			Group X			Group Y		
	Black	White	Difference	Black	White	Difference	Black	White	Difference
English	26.4	73.6	47.2	77.2	22.8	54.4	94.0	6.0	88.0
Social Science	31.5	68.5	37.0	74.4	25.6	48.8	94.9	5.1	89.8
Mathematics	17.4	82.6	65.2	72.9	27.1	45.8	93.5	6.5	87.0
Science	13.3	86.7	73.4	73.8	26.2	47.6	93.3	6.7	86.6
TOTAL	22.2	77.9	55.7	74.6	25.4	49.2	93.9	6.1	87.9

how fast this type of planning is extended to other grade levels.

Matzen (1965) studied a total of 1,100 black and white students in grades 5 and 7 in 11 different schools in the San Francisco Bay area to determine the relationship of the proportion of black children in a classroom to the mean scholastic achievement of black and white students. Test findings showed a tendency for both achievement and IQ to vary inversely with percent of black students, with, however, numerous exceptions. Achievement varied directly with socioeconomic level; when IQ and socioeconomic status were held constant, achievement tended to fall as the percent of black students rose, but the tendency was not strong enough to reach statistical significance. In the fifth grade, where students were less homogeneously grouped than in the seventh grade, the black-white differentials in achievement were greater. In the seventh grade, with bright black children and bright white children in the same classrooms, black-white differences in achievement were minimized. Matzen's findings may be interpreted in many ways, but it is perhaps best to note that they are consonant with those of McPartland's more substantial study discussed below.

Two significant analyses of data from the Coleman Report (1966) are extremely pertinent to any current discussion of the impact of ability grouping on school achievement of minority groups. The first of these is the work of McPartland (1968, 1969), a colleague of Coleman's on the Educational Opportunities Survey, which resulted in the most comprehensive body of data ever collected on public schools and their students in the United States. The second is by Mayeske (1970), charged with colleagues at the U.S. Office of Education with the responsibility of illustrating and documenting the structure and functioning of the American public school system.

McPartland (1968, 1969) analyzed data on students from a sample of schools selected from metropolitan areas of the New England and Middle Atlantic states participating in the Survey. He studied 5,075 ninth-grade black students who had attended their present schools in the previous years, using three variables to set up cross classification: a six-level family background scale constructed from students' reports of their mothers' education and students' responses on a nine-item check list of possessions in the home; the percent of white students in the ninth grade of a student's school, partitioned into four categories; and four groupings derived from the student's report of the proportion of his classmates who were white. Average achievement scores on a 60-item test of verbal ability derived from the *School and College Ability Test* were calculated within cells of cross-classification of the variables used. Summary measures were then derived from McPartland's cross tabulations. From the analysis of ninth-grade students in the metropolitan Northeast, McPartland concluded that the potential favorable effects of school desegregation on black achievement can be offset by segregation within the school. He found that only black students in mostly white classes demonstrate any added achievement growth due to attendance at mostly white schools. On the other hand, he found, class desegregation has a favorable effect on black student verbal achievement, no matter what the racial enrollment of the school. He provides evidence that the differences in verbal achievement between black students in mostly white classes and black students in mostly black classes cannot be explained by selection processes which operate within a given school.

The information collected from students in the Coleman study concerned (a) the students' programs of study, (b) the particular courses in which students were enrolled, and (c) the track levels to which students were assigned in their English classes. It is clear from McPartland's analysis that within schools of similar racial composition the program of study in

which a student is enrolled has a strong influence on the chance that he will be in a majority white class. Generally, students enrolled in the college preparatory program are most likely to be in classes which are more than 50 percent white. Conversely, students in vocational, commercial, or industrial arts programs are least likely to have mostly white classmates. McPartland points out that the schools which are exceptions to this generalization are those where only a small fraction of the student body is white. The reason for this is that, in contrast to most other schools, "the white students in many of these predominantly black schools are among the poorest students in the school." Therefore, except for predominantly Negro schools with a few white students, the practical consequence of program assignments within schools on the racial composition of a Negro student's classes is the same. Students who tend to achieve in academic areas, as measured by various reading and arithmetic achievement tests, tend to be selected or enrolled in advanced academic programs which tend to have more white students than in non-academic courses of study.

McPartland presents additional data which highlight the relation between program of study and classroom racial composition. These illustrate that within schools of similar racial composition, black children in mostly white classes are most frequently enrolled in academic courses, and least likely to be taking vocational, commercial, industrial arts, or home economics courses. Says McPartland:

> The most dramatic positive differences with the fewest reversals are for courses which are likely to be part of a college preparatory program rather than some other program: the science and foreign language courses. But even for the course work likely to be required for most students, such as English and mathematics, there is some evidence that enrollment in these subjects is related to the racial composition of a Negro student's classmates. It is with courses such as mathematics and English that separate classes will be organized according to the achievement level of students to be assigned to the class.

Also, with respect to the racial composition of *classes* as a direct result of tracking or ability grouping, McPartland documents that the largest proportion of the students in the highest track have mostly white classmates. That is, half of all black children in the high English track have more than half white classmates in schools which enroll 50 to 69 percent whites, while approximately 33 percent of the Negro students in the middle and lowest tracks are in such classes.

Finally, McPartland goes on to show, that this separation of pupils ethnically has an effect on achievement of the Negro students. Carefully controlling for home background factors, he shows that only when a majority of classmates of black students are from the predominant white group do the Negro students show benefits from desegregation. It is the improved learning of these black students that makes Negro achievement in desegregated schools improve on the average; students in other classes show no improvement and even possibly slight loss.

Mayeske (1970) reported further data from the Coleman Report that are especially pertinent here. In his analysis of the data, Mayeske found a relationship at the first-grade level between achievement levels of entering students and the attributes of the schools they attended. Schools with entering students of higher levels of achievement had associated with them teachers who possessed higher verbal skills, who tended to be white, and who expressed a preference for working with high ability students. He found that these relationships with achievement tended to increase at the higher grade levels. The same was true of the relationship of achievement with the students' social background.

Mayeske refers to this phenomenon as the "ecological-functional dilemma." At the beginning of the first grade, students tend to be allocated into schools on the basis of their social backgrounds. Certain relationships, which Mayeske refers to as "ecological relationships," are observed between the attributes of the students and their schools. Over time, since students with high social backgrounds benefit more from their schooling, ecology and the school's influence become more and more intertwined so that it becomes increasingly difficult to separate out their independent influences. The schools reflect the deep-seated social problem of ethnic separation which permeates almost every aspect of American life. This basic problem, according to Mayeske, in the main is that a person's birth into a particular stratum of society plays a large role in determining where that individual will go and will not go in the scheme of things. The problem is made even more difficult because one's skin color and language habits tend to be associated with one's position within the social structure. If Mayeske's interpretation has any validity, the schools alone cannot rectify the problem, although they can play an ameliorative role; the problem must be attacked on a number of different fronts, such as jobs, housing, schooling, and various other areas characterized by separation and segregation.

Mayeske concludes, as did Coleman, that the schools play an important role in promoting achievement for all students; but, as the schools are currently constituted, students from the higher socioeconomic levels, of whom most are white, benefit more from

attending school than students from the lower socioeconomic strata, many of whom are non-white. He suggests that to break these socioeconomic background barriers, innovations that differ radically from past practices might be tried in situations so structured that the results of the innovations can be clearly demonstrated. Some suggested innovations include more socioeconomically and racially balanced student bodies and teaching staffs, competitive school systems or voucher systems whereby the student and his family can select services from a variety of sources, and concern by real estate people with the improvement of the quality and composition of schools rather than with the maintenance of racially segregated communities in terms of available housing.

Finally, Maynor (1970) compared achievement of 680 black, 127 Indian, and 608 white students before and after the first year of integration in grades 6 through 12 in Hoke County, North Carolina. The slopes of the regression lines for achievement, relative to grade placement, on the reading, language, mathematics, and total scores on the *California Achievement Tests* showed no change, so it was possible to compare differences in average achievement over the range of grades. Blacks showed gains in all parts, but only those in mathematics and total scores were significant at the 5 percent level. Indians and whites showed neither gains nor losses. Blacks did their best when taught by Indian teachers.

NON-NEGRO MINORITIES

Many of the educational disabilities which burden Negro Americans are shared by Mexican Americans, Puerto Ricans, and Indian Americans. Weinberg (1970) goes so far as to say that these three minority groups are the most educationally disadvantaged in the United States.

The urban Negro ghetto is reenacted in Mexican-American neighborhoods in the cities of southern California and the Southwestern states; the Puerto Rican communities in New York and other cities of the Northeast are as isolated from the white communities as is Harlem; and the Indian Americans, especially those living on or near reservations, are the most segregated of all. In recent years, a fourth minority group, the expatriate Cubans in the Southeastern states, especially Florida, have become groups alone.

Belonging to an ethnic minority in the United States and being poor besides creates a common plight for all these people. For Mexican Americans and Puerto Ricans—and, more recently, the Cubans—a "foreign" language has become a barrier to normal educational progress. The exclusive use in most schools of English as the language of instruction, among children understanding this language little or not at all, by teachers not knowing Spanish, has created multiple problems. Add to this the lack of sensitivity on the part of teachers to sociocultural differences in children, and an almost intolerable situation exists in the schools.

Weinberg devotes an entire chapter in his *Desegregation Research: An Appraisal* to summarizing research studies of the past 35 years devoted to the exploration of the problems of these minority groups. The research findings are similar to those reported earlier for Negro students and for both black and white students of low socioeconomic status. On the whole, children of the non-Negro minority groups compare unfavorably with middle-class white children with respect to IQ and academic achievement level, segregation has been their usual lot in school, they consider themselves to be inferior to the majority whites, and their educational and occupational aspirations are likely to be low. As with the Negro, in those schools in which ability grouping is practiced, classes almost homogeneous racially have been created. And as with the Negro also, the greater the degree of contact the minority child has with the white man's culture, the higher he scores on educational tests, and the greater his progress academically, the more favorable his self-concept, and the higher his aspirations.

Carter (1970) has described in detail the history of the educational neglect of Mexican-American children. While there are some exceptions, the majority of Mexican Americans have lower-class status. Even though the children may attend mixed schools, in reality they may be isolated from their Anglo and middle-class Mexican-American peers. School policy and practice have contributed to this isolation, tending to reinforce the ethnic and social clevage that exists in the Southwest. The school reflects the community and tends to perpetuate the separation of Mexican and Anglo roles and aspirations.

Special compensatory programs for Mexican-American children are becoming almost universal in Southwestern schools. Compensatory classes requiring attendance for part of the day are most frequently encountered; this kind of program does not isolate the children to an unwarranted degree. When compensatory programs require fulltime attendance, the Mexican-American children are substantially isolated, in essence attending, within an ethnically mixed institution, a subschool from which they cannot break out.

According to Carter, rigid ability grouping, or tracking, in one form or another is widely practiced in Southwestern schools. Appraisal of intellectual capacity and academic achievement, whether by standardized tests or other means, usually determines track assignment. Since Mexican-American children, espe-

cially those of low socioeconomic status, tend to fall below school or national norms, they are greatly overrepresented in the lower-ability tracks while the Anglos are overrepresented in the middle- and high-ability tracks. Although a first grader has a better chance to change tracks than a tenth grader, once a student is tracked at any level, movement upward is difficult.

Little research concerning the effects of tracking on the achievement and attitudes of Mexican-American students has been done. Regardless of the effects on achievement, however, Carter contends that the track system adversely affects both teachers' and students' expectations and their subsequent behavior. Since it unduly isolates Mexican-American youth from equal-status interaction with others, it maintains cultural differences and slows down the process of acculturation.

Carter writes that the information collected concerning the practice of tracking in the Washington, D.C., schools at the time of the Hobson vs. Hansen case could equally well describe the practice in most Southwestern schools. To what degree the impact of the Court decision in the Hobson vs. Hensen case may influence Mexican-American organizations to attempt legal recourse to obtain equal educational opportunities for Mexican-American children is a matter of conjecture at the present time.

In response to a request for information about grouping practices based on test scores and school problems they might present for American Indian children, Havighurst (1970) offered this information based on the National Study of American Indian Education:

> ... most Indian children are in schools where they are in the majority. In these schools, most of which are relatively small, there is seldom any ability grouping.
>
> Another category of Indian student consists of those who live near an Indian reservation but attend ... a high school that has a majority of non-Indian students (for example, Cutbank, Montana; Moclips, Washington; Gallup and Albuquerque, New Mexico; Globe, Arizona). In these communities the Indians generally perform below the average of the non-Indian. However, there is not much grouping in these communities, which are generally rather small in their school populations.
>
> A third category consists of Indian students in relatively large urban centers where the Indians seldom go above 10 percent in any one school and often are present in less than one percent proportions. Here there may be some ability grouping based on tests and depending on the policy of the school system. Almost all of the big cities from Chicago on West have these kinds of Indian minorities. Also, you find them in smaller urban centers like Mesa, Arizona; Bell Gardens, Los Angeles; Tucson, Arizona. At the high school level we find that there is some ability grouping based on tests in a number of high schools. Generally, the Indian youngsters tend to be placed in the average or below average ability groups. Still there are usually a few who do well on tests and get placed in the higher ability groups.

It would appear from Havighurst's letter that American Indian children are less generally affected than children of other minority groups by ability grouping practices. Certainly there are no situations in which they are isolated from the white majority as a result of ability grouping. The reader may wish to refer to the study of Maynor discussed briefly on page 52 of this section.

In summary, the reported information about non-Negro minorities is scant, but consonant with the findings for Negro students. The special connotation of "language handicap" for Spanish-speaking or bilingual minorities in the United States could be studied in terms of test results, but is more properly seen in the broader context of pluralistic education, needed respect for minority cultures, and humanitarian concern for all children on an equal basis of acceptance and assistance as well as opportunity.

SUMMARY AND CONCLUDING REMARKS

This second section summarizes, in as readable format as we could devise, the important studies relevant to *The Impact of Ability Grouping on School Achievement, Affective Development, Ethnic Separation and Socioeconomic Separation.* It is supported in detail by an extensive bibliography of historical and timely references. The reader may expect to find here sufficient discussion of major findings and enough illustrative material to clarify the points made. Careful perusal of the references will allow the reader to fill in the greater detail he may desire at any point without our having slowed other readers not interested in so much detail about that point. On the other hand, we would suggest that what is presented here will be merely supported, clarified, or expanded, but not contradicted in any essential respect by reading the references. Nor do we feel that we have omitted relevant references. So far as we could make it, then, this is a summary and guide to the essential truth about this topic.

We are concerned here with schemes of organization of schools into classroom groups on the basis of test results or judgments relative to the ability of students, in such a way as to bring together in instructional groups children of a given age or grade who are most nearly equal in relevant abilities. Grouping and

regrouping within the classroom for instruction of those needing assistance in mastering particular bits of skill or information is considered a normal and desirable instructional practice.

Briefly, we find that ability grouping as defined above shows no consistent positive value for helping students generally, or particular groups of students, to learn better. Taking all studies into account, the balance of findings is chiefly of no strong effect either favorable or unfavorable. Among the studies showing significant effects, the slight preponderance of evidence showing the practice favorable for the learning of high ability students is more than offset by evidence of unfavorable effects on the learning of average and low ability groups, particularly the latter. There is no appreciable difference in the effects at elementary and secondary school levels. Finally, those instances of special benefit under ability grouping have generally involved substantial modification of materials and methods, which may well be the influential factors wholly apart from grouping.

The findings regarding impact of ability grouping on the affective development of children are essentially unfavorable. Whatever the practice does to build (inflate?) the egos of children in the high groups is overbalanced by evidence of unfavorable effects of stigmatizing average and low groups as inferior and incapable of learning.

In the absence of evidence of positive effects on learning and personal development of children, and in the light of negative effects on the scholastic achievement and self-concepts of low ability groups, the tendency of ability grouping to separate children along ethnic and socioeconomic lines must be deemed to discriminate against children from low socioeconomic classes and minority groups. The mechanism may be said to operate primarily by denying the low groups the scholastic stimulation of their more able peers, and by stigmatizing the low groups as inferior and incapable of learning in their own eyes and those of their teachers. McPartland's data are particularly significant in showing that whatever superior achievement is shown by blacks in desegregated schools, is produced by the superior achievement of blacks in predominantly white (middle class) classroom groups.

Throughout this document we have moved back and forth between ethnic and socioeconomic variables. The fundamental fact of the situation is that minority group membership is consistently and strongly associated with low socioeconomic status. Conversely, high socioeconomic status is strongly associated with membership in the predominant "white" culture. It has not seemed practical or profitable to attempt to delineate these effects differentially. The practical circumstance is that minority groups preponderantly suffer the disadvantages of low socioeconomic status, increased by the fact of being more immediately identifiable by physical appearance. One can only hope that continuing attention will be given to the socioeconomic factor as basic.

Four brief footnotes. First, ability grouping is undesirable even where ethnic and socioeconomic factors are not present, as they generally are. Second, removal of ability grouping has no effect on ethnic discrimination where population movement has already produced ethnic isolation. Third, studies of other minority groups than blacks are needed to bring proper attention to the plight of these smaller minority groups, whose present situation is quite as serious, but not as prominent. Fourth, socioeconomic isolation needs to be elevated to central attention.

Finally, nothing included here may be taken as conclusive evidence that a plan of classroom organization and related procedures may not be effective if well designed to achieve its purpose—for gifted, for mentally retarded, or for children generally. The evidence simply indicates that ability grouping *per se* tends to be ineffective and do more harm than good. Any procedure that involves ability grouping and corollary ethnic separation must be justified in terms of other strong evidence of likely beneficial effects.

REFERENCES

Adkison, Marion Robert. "A Comparative Study of Pupil Attitudes under Conditions of Ability and of Heterogeneous Grouping," Claremont, California: Claremont Graduate School, 1964. Abstract: *Dissertation Abstracts* 28:3869-A, April 1968.

Antonovsky, Aaron and M.J. Lerner. "Negro and White School Youth in Elmira." In *Discrimination and Low Income,* Aaron Antonovsky and Lewis L. Lorwin, eds. New York: New School for Social Research, 1959. pp. 103-45.

Bacher, Jesse Hiram. "The Effect of Special Class Placement on the Self-Concept, Social Adjustment, and Reading Growth of Slower Learners." New York: New York University, 1964. Abstract: *Dissertation Abstracts* 25:7071, June 1965.

Barber, Theodore X. and David S. Calverley, Albert Forgione, John D. McPeake, John F. Chaves, and Barbara Bowen. "Five Attempts to Replicate the Experimenter Bias Effect." *Journal of Consulting and Clinical Psychology* 33(1):1-6, February 1969.

Barker Lunn, Joan C. *Streaming in the Primary School.* London: National Foundation for Educational Research in England and Wales, 1970.

Bernstein, S. and D. Esposito. "On Grouping in the Experimental Elementary School Project," paper prepared for ERIC Information Retrieval Center for the Disadvantaged, Teachers College, Columbia University, November 1969.

Billett, R. O. *The Administration and Supervision of Homogeneous Groupings.* Columbus, Ohio: Ohio State University, 1932.

Borg, Walter R. *Ability Grouping in the Public Schools,* Second Edition. Madison, Wisconsin: Dembar Educational Research Services, Inc., 1966.

Borg, Walter R. and Tony Pepich. "Grouping of Slow Learning High School Pupils." *Journal of Secondary Education,* 41:231-38, May 1966.

Borg, Walter R. and M. Richard Maxfield. "A Study of the Effects of Different Kinds of Ability Grouping on Personal Relationships Among High School Students." Washington. D.C.: U. S. Office of Education, Bureau of Research, Project 5-8042, December 1967. Abstract: *Research in Education* ED 017 002, 1968.

Bouri, Janet and Joan Barker Lunn. *Too Small to Stream.* A Study of Grouping in Small Junior High Schools. Occasional Publication Series No. 21. London: National Foundation for Educational Research in England and Wales, 1969.

Brophy, Jere E. and T. L. Good. "Teachers' Communication of Differential Expectations for Children's Classroom Performance: Some Behavioral Data," Report Series No. 25. Austin, Texas: Research and Development Center in Teacher Education, University of Texas, 1969 (Also, in press, *Journal of Educational Psychology,* 1970).

Byers, Loretta. "Ability Grouping—Help or Hindrance to Social and Emotional Growth." *School Review,* 69:449-56, Winter 1961.

Carter, Thomas P. *Mexican Americans in School: A History of Educational Neglect.* New York: College Entrance Examination Board, 1970.

Clark, Kenneth B. "Educational Stimulation of Racially Disadvantaged Children." In *Education in Depressed Areas,* A. Harry Passow, ed. New York: Bureau of Publications, Teachers College, Columbia University, 1963.

Coleman, Hubert A. "The Relationship of Socio-economic Status to the Performance of Junior High School Students." *Journal of Experimental Education,* 9:61-63, September 1940.

Coleman, James S. and Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. *Equality of Educational Opportunity.* Washington, D.C.: Government Printing Office, 1966.

Cox, F. N. "Educational Streaming and General and Test Anxiety." *Child Development,* 33:381-90, June 1962.

DeGrow, Gerald S. "A Study of the Effects of the Use of Vertical Reading Ability Grouping for Reading Classes as Compared with Heterogeneous Groupings in Grades Four, Five, and Six in the Port Huron Area Public Schools of Michigan over a Three-Year Period." Ann Arbor, Michigan: University of Michigan, 1963. Abstract: *Dissertation Abstracts* 24:3166, February 1964.

Deitrich, Francis R. "Comparison of Sociometric Patterns of Sixth-Grade Pupils in Two School Systems: Ability Grouping Compared with Heterogeneous Grouping." *Journal of Educational Research,* 57:507-13, July 1964.

Deutsch, Martin and Bert Brown. "Social Influences in Negro-White Intelligence Differences." *The Journal of Social Issues,* 20:24-36, April 1964.

Dreger, R. M. and K. S. Miller. "Comparative Psychological Studies of Negroes and Whites in the United States." *Psychological Bulletin.* 57:361-402, September 1960.

Drews, E. M. *Student Abilities, Grouping Patterns and Classroom Interactions.* East Lansing, Michigan: Office of Research and Publications, Michigan State University, 1963.

Dyson, Ernest. "A Study of the Relationships Between Acceptance of Self, Academic Self-Concept, and Two Types of Grouping Procedures Used with Seventh Grade Pupils." Philadelphia, Pennsylvania: Temple University, 1965. Abstract: *Dissertation Abstracts* 26:1475-76, September 1965.

Eash, Maurice J. "Grouping: What Have We Learned?" *Educational Leadership,* 18:429-34, April 1961.

Ekstrom, Ruth B. *Experimental Studies of Homogeneous Grouping: A Review of the Literature.* Princeton, New Jersey: Educational Testing Service, 1959.

Fick, Wayne W. "The Effectiveness of Ability Grouping in Seventh Grade Core Classes." Lawrence, Kansas: University of Kansas, 1962. Abstract: *Dissertation Abstracts* 23:2753, February 1963.

Flanagan, J. C. and F. B. Davis, J. T. Dailey, M. F. Shaycoft, D. B. Orr, I. Goldberg, and C. A. Neyman, Jr. *The American High-School Student.* Final Report, Cooperative Research Project No. 635, U.S. Office of Education. Pittsburgh, Pennsylvania: Project Talent Office, University of Pittsburgh, 1964.

Flowers, Charles Edward. "Effects of an Arbitrary Accelerated Group Placement on the Tested Academic Achievement of Educationally Disadvantaged

Students." New York: Columbia University, 1966. Abstract: *Dissertation Abstracts* 27:991-A, October 1966.

Franseth, Jane. *A Guide to Research and Informed Judgment on Grouping Children,* Education Briefs No. 40. Washington, D.C.: Office of Education, U.S. Department of Health, Education, and Welfare, May 1964.

Goldberg, Miriam L. "Factors Affecting Educational Attainment in Depressed Areas." In *Education in Depressed Areas,* A. Harry Passow, ed. New York: Bureau of Publications, Teachers College, Columbia University, 1963. pp. 68-99.

Goldberg, Miriam L. and A. Harry Passow and Joseph Justman. *The Effects of Ability Grouping.* New York: Teachers College Press, Columbia University, 1966.

Goldstein, Bernard. *Low Income Youth in Urban Areas.* A Critical Review of the Literature. New York: Holt, Rinehart, and Winston, Inc., 1967.

Goldworth, Mary. "The Effects of an Elementary School Fast-Learner Program on Children's Social Relationships." *Exceptional Children,* 26:59-63, October 1959.

Good, T. L. and Jere E. Brophy. "Do Boys and Girls Receive Equal Opportunity in the First Grade Reading Instruction?" Research Series No. 24. Austin, Texas: Research and Development Center in Teacher Education, University of Texas, 1969.

Goodlad, John I. "Classroom Organization." In *Encyclopedia of Educational Research,* Third Edition, Chester W. Harris, ed. New York: The Macmillan Company, 1960. pp. 221-25.

Griffin, A. "Effects of Secondary School Organization on the Development of Intellectual Attainments in England and Attitudes to School." *British Journal of Educational Psychology,* 39:191, June 1969.

Havighurst, Robert J. Information from the Study of American Indian Education contained in a letter to the project committee, dated February 13, 1970.

Heathers, Glen. "Grouping." In *Encyclopedia of Educational Research,* Fourth Edition, Robert L. Ebel, ed. New York: The Macmillan Company, 1969. pp. 559-70.

Hobson v. Hansen. Congressional Record, June 21, 1967.

Justman, Joseph. *A Comparison of the Functioning of Intellectually Gifted Children Enrolled in Special Progress and Normal Progress Classes in Junior High School.* New York: Columbia University, 1953. Abstract: *Dissertation Abstracts* 14:65, January-April 1954.

Kariger, Roger H. "The Relationship of Lane Grouping to the Socioeconomic Status of the Parents of Seventh-Grade Pupils in Three Junior High Schools." East Lansing, Michigan: Michigan State University, 1962. Abstract: *Dissertation Abstracts* 23:4586, June 1963.

Kellmer-Pringle, M. L. and T. Cox. "The Influence of Schooling and Sex on Test and General Anxiety as Measured by Sarason's Scales." *Journal of Child Psychology, Psychiatry,* 4:157-65, December 1963.

Kennedy, Wallace A. and Vernon Van De Riet and James C. White, Jr. "A Normative Sample of Intelligence and Achievement of Negro Elementary School Children in the Southeastern United States." *Monographs of the Society for Research in Child Development,* 28:6, 1963.

Kline, Rosemary Elizabeth. "A Longitudinal Study of the Effectiveness of the Track Plan in the Secondary Schools of a Metropolitan Community." St. Louis, Missouri: St. Louis University, 1963. Abstract: *Dissertation Abstracts* 25:324, July 1964.

Levy, Philip and Stan Gooch and M. L. Kellmer-Pringle. "A Longitudinal Study of the Relationship between Anxiety and Streaming in a Progressive and a Traditional Junior School." *British Journal of Educational Psychology,* 39:166-73, June 1969.

Locke, Carolyn Jane. "The Effect of Two Plans of Elementary School Organization on Rapid Learners and Non-Rapid Learners." Denton, Texas: North Texas State University, 1962. Abstract: *Dissertation Abstracts* 23:1289, October 1962.

Loomer, Bradley Max. "Ability Grouping and Its Effect Upon Individual Achievement." Iowa City, Iowa: State University of Iowa, 1962. Abstract: *Dissertation Abstracts* 23:1581, November 1962.

Luchins, Abraham S. and Edith H. Luchins. "Children's Attitudes Toward Homogeneous Groupings." *Pedagogical Seminary and Journal of Genetic Psychology,* 72:3-9, March 1948.

Lutrell, Jack. "A Comparative Investigation of the Academic Achievement and Personality Development of Gifted Sixth-Grade Pupils in a Special Class and in Regular Classrooms in the Public Schools of Greensboro, North Carolina." Chapel Hill, North Carolina: University of North Carolina, 1958. Abstract: *Dissertation Abstracts* 19:2536, April-June 1959.

Mann, Horace. "How Real Are Friendships of Gifted and Typical Children in a Program of Partial Segregation?" *Exceptional Children,* 23:199-201, 206, February 1957.

Mann, Maxine. "What Does Ability Grouping Do to the Self-Concept?" *Childhood Education,* 36:356-60, April 1960.

Matzen, Stanley P. "The Relationship between Racial Composition and Scholastic Achievement in Elementary School Classrooms." Palo Alto, California: Stanford University, 1965. Abstract: *Dissertation Abstracts* 26:6475, May 1966.

Mayeske, George and Carol E. Wisler, Albert E. Beaton, Jr., Frederic Weinfeld, Wallace M. Cohern, Tetsuo Okoda, John M. Proshek, and Kenneth S. Tabler. *A Study of Our Nation's Schools.* Washington, D.C.: Government Printing Office, 1969.

Maynor, Waltz. "Academic Performance and School Integration: A Multi-Ethnic Analysis." Unpublished dissertation, Duke University, 1970.

McCandless, B. R. *Children: Behavior and Development.* New York: Holt, Rinehart and Winston, Inc., 1967.

McDill, Mary S. and Arthur L. Stinchcombe and Dollie Walker. "Segregation and Educational Disadvantage: Estimates of the Influence of Different Segregating Factors." *Sociology of Education,* 41:239-46, Summer 1968.

McPartland, James. "The Relative Influence of School Desegregation and of Classroom Desegregation on the Academic Achievement of Ninth-Grade Negro Students." *Journal of Social Issues,* 25(3):93-102, Summer 1969.

McPartland, James. *The Segregated Student in Desegregated Schools. Sources of Influence on Negro Secondary Students.* Baltimore, Maryland: Johns Hopkins University, June 1968.

Mehl, Robert Franklin, Jr. "A Study of Relationships between Homogeneous Grouping in the School and the Social Class Structure in an Up-State New York Community." Albany, New York: State University of New York at Albany, 1965. Abstract: *Dissertation Abstracts* 27:4085-A, June 1967.

Mikkelson, James Eliot. "An Experimental Study of Selective Grouping and Acceleration in Junior High School Mathematics." St. Paul, Minnesota: University of Minnesota, 1962. Abstract: *Dissertation Abstracts* 23:4226, May 1963.

Miner, B. "Sociological Background Variables Affecting School Achievement." *Journal of Educational Research,* 61:372-81, April 1968.

Morgenstern, Anne. "A Comparison of the Effects of Heterogeneous and Homogeneous (Ability) Grouping on the Academic Achievement and Personal-Social Adjustment of Sixth-Grade Children." New York: New York University, 1963. Abstract: *Psychological Abstracts* 24:1054, September 1963.

National Education Association, Research Division. *Ability Grouping,* Research Summary 1968-S3. Washington, D.C.: National Education Association, 1968.

New York State Education Department, Division of Research. *Racial and Social Isolation in the Schools* (Preliminary Report). Albany, New York: Office of Public Information, New York State Education Department, 1969.

Ogletree, Earl. "Homogeneous Ability Grouping—British Style." *Peabody Journal of Education,* 47:20-25, July 1969.

Olivarri, Martin C. "Some Relationships of Ability Grouping to Student Self-Concept." Berkeley, California: University of California, 1967. Abstract: *Dissertation Abstracts* 28:2518A, January 1968.

Otto, Henry. "Classification of Pupils." In *Encyclopedia of Educational Research,* Second Edition, Walter S. Monroe, ed. New York: The Macmillan Company, 1950. pp. 376-78.

Otto, Henry J. "Elementary Education—II, Organization and Administration." In *Encyclopedia of Educational Research,* First Edition, Walter S. Monroe, ed. New York: The Macmillan Company, 1941. pp. 428-46.

Passow, A. Harry. *Toward Creating a Model Urban School System: A Study of the Washington, D.C. Public Schools.* New York: Teachers College, Columbia University, 1967.

Peterson, Richard Lee. "An Experimental Study of the Effects of Ability Grouping in Grades Seven and Eight." St. Paul, Minnesota: University of Minnesota, 1966. Abstract: *Dissertation Abstracts* 28:103-A, July 1967.

Plainfield Study. *Grouping Students for Instruction in the Plainfield, New Jersey, School System.* New York: Institute of Field Studies, Teachers College, Columbia University, July 1969.

Provus, Malcolm M. "Ability Grouping in Arithmetic." *Elementary School Journal.* 60:391-98, April 1960.

Rosenthal, Robert. "Empirical vs. Decreed Validation of Clocks and Tests." *American Educational Research Journal,* 6:689-91, November 1969.

Rosenthal, Robert and Lenore Jacobson. *Pygmalion in the Classroom.* New York: Holt, Rinehart and Winston, Inc., 1968.

Rudd, W. G. "The Psychological Effect of Streaming by Attainment." *British Journal of Educational Psychology,* 28:47-60, February 1958.

Sarthory, Joseph A. "The Effects of Ability Grouping in the Multi-Cultural School Situation." Albuquerque, New Mexico: University of New Mexico, 1968. Abstract: *Dissertation Abstracts* 29:451-A, August 1968.

Simpson, R. E. and Ruth A. Martinson. *Educational Programs for Gifted Pupils.* Sacramento, California: California State Department of Education, January 1961.

Stinchcombe, Arthur L. and Mary S. McDill and Dollie R. Walker. "Demography of Organizations." *American Journal of Sociology,* 74:221-29, November 1968.

Stinchcombe, Arthur L. and Mary S. McDill and Dollie R. Walker. "Is There a Racial Tipping Point in Changing Schools?" *Journal of Social Issues* 25(1):127-36, Winter 1969.

Thorndike, Robert L. "But You Have to Know How to Tell Time." *American Educational Research Journal* 6:692, November 1969.

Thorndike, Robert L. "Review of Pygmalion in the Classroom." *American Educational Research Journal* 5:708-711, November 1968.

Tobin, John Francis. "An Eight-Year Study of Classes Grouped Within Grade Levels on the Basis of Reading Ability." Boston, Massachusetts: Boston University, 1965. Abstract: *Dissertation Abstracts* 26:5141, March 1966.

Torgelson, John Winston. "A Comparison of Homogeneous Grouping for Below-Average Junior High School Students." Minneapolis, Minnesota: University of Minnesota, 1963. Abstract: *Dissertation Abstracts* 25:2300-01, October 1964.

U. S. Commission on Civil Rights. *Racial Isolation in the Public Schools,* I and II. Washington, D. C.: Government Printing Office, 1967.

U. S. National Advisory Commission on Civil Disorders. *Report of the National Advisory Committee on Civil Disorders.* Washington, D.C.: Government Printing Office, 1968.

Walker, Dollie R. and Arthur L. Stinchcombe and Mary S. McDill. *School Desegregation in Baltimore.* Baltimore, Maryland: Center for the Study of Social Organization of Schools, John Hopkins University, August 1967.

Weinberg, Meyer. *Desegregation Approach: An Appraisal,* Second Edition. Bloomington, Indiana: Phi Delta Kappa, May 1970.

Wilcox, John. "A Search for the Multiple Effects of Grouping Upon the Growth and Behavior of Junior High School Pupils." Ithaca, New York: Cornell University, 1963. Abstracts: *Dissertation Abstracts* 24:205, July 1963.

Wilhelms, Fred T. and Dorothy Westby-Gibson. "Grouping: Research Offers Leads." *Educational Leadership,* 18:410-13, 476, April 1961.

Willcutt, Robert Ernest. "Ability Grouping by Content Subject Areas in Junior High School Mathematics." Bloomington, Indiana: Indiana University, 1967. Abstract: *Dissertation Abstracts* 28:2152A, December 1967.

Willig, C. J. "Social Implications of Streaming in the Junior School." *Educational Research,* 5:151-54; February 1963.

Wilson, Alan B. "Educational Consequences of Segregation in a California Community." In *Racial Isolation in the Public Schools* II. A Report of the U.S. Commission on Civil Rights. Washington, D.C.: Government Printing Office, 1967. pp. 165-206.

Zweibelson, I. and M. Bahmuller and L. Lyman. "Team Teaching and Flexible Grouping in the Junior High School Social Studies." *The Journal of Experimental Education,* 34:20-32; Fall 1965.

III. THE PROBLEMS AND UTILITIES INVOLVED IN THE USE OF TESTS FOR GROUPING CHILDREN WITH LIMITED BACKGROUNDS

The search for useful information regarding the validity and reliability of standardized aptitude and achievement tests for use in grouping children with limited backgrounds for purposes of instruction has been an exhaustive but, unfortunately, not a very productive one. Not a single study, for example, among the more than two hundred located was found to involve all three aspects of the topic: test validity and reliability, culturally limited populations, and homogeneous grouping. It has been necessary, therefore, to attempt to go beyond the data presented and to make calculated inferences as to what might be expected to occur under certain combinations of circumstances.

DEFINITION OF TERMS

The definition of a few terms is in order here if the intent of this section is to be clearly understood. These definitions may be read first or in conjunction with the discussion that follows. They are presented in a sequence of importance for understanding the material of the section. Wherever a term used in a definition is not understood, its definition is to be found later on.

1. In this section, concern will be for the *validity* not only of the tests themselves but also of their use for the whole population. Are the tests giving us the kind of information about students and about programs of instruction that we really want to know? In particular, do the tests provide comparable information about students with different backgrounds that can be useful in conducting the instructional program? Note particularly the definition of *construct* or *pure validity* given last.

The *validity* of a test refers to the extent to which a test does the job for which it is intended. Validity has different connotations for various kinds of tests and, accordingly, different kinds of validity are appropriate for them. For example, the validity of an achievement test is the extent to which the content of the test represents a balanced and adequate sampling of the outcomes (knowledge, skills, etc.) of the course or instructional program it is intended to cover (*content, face,* or *curricular validity*). The validity of an aptitude or readiness test is the extent to which it accurately indicates future learning success in the area for which it is used as a predictor *(predictive validity).* The validity of a personality test is the extent to which the test yields an accurate description of an individual's personality traits or personality organization as of that moment (*status* or *concurrent validity*).

The *validity* of a test or of a procedure for the use of a test for a particular purpose involves a combination of concurrent validity for indicating the present status of individuals in mastering a subject, predictive validity for indicating the probable later achievement of individuals in mastering that subject under specified instructional procedures, and freedom from correlation with extraneous variables on the part of the original or final measures of achievement. This total requirement may be called *construct* or *pure validity.* This concept of validity may be extended to other measures—self-concept ratings, personality measures, etc.—by substituting such terms for "test" in this definition.

2. The *reliability* of a test refers to the extent to which a test is consistent in measuring whatever it does measure: dependability, stability, relative freedom from errors of measurement. It is usually estimated by some form of *reliability coefficient* or by the *standard error of measurement.* The higher the *reliability coefficient* and the smaller the *standard error of measurement,* the more reliable is the test.

Reliability coefficients take their names from the method of determination. In this section we will be most frequently concerned with the *alternative form coefficient,* which is generally obtained by giving two parallel forms of a test (with equal content, means, and variances) to the same group of individuals on closely succeeding days and correlating the results; the *split-half coefficient,* which is obtained by correlating scores on one half of a test with scores on the other half; the *Kuder-Richardson coefficient,* which is obtained from item statistics of a single administration of one form of a test; and the *test-retest coefficient,* which is obtained by administering the same test a second time after a short interval and correlating the two sets of scores. The alternate form estimate is generally preferred because it reflects the day-to-day variability implicit in ordinary use of tests.

3. The *standard error of measurement* is an estimate of the magnitude of the "error of measurement" in a score—the amount by which an obtained score differs from a hypothetical true score. It is the standard deviation of the differences between actual scores and theoretical true scores of the same individuals on a test. The standard error is an amount such that in about two thirds of the cases the obtained score would not differ from the true score by more than one standard error.

4. A *standard deviation* is a measure of the variability or dispersion of a set of scores. The more the scores cluster around the mean, the smaller the standard deviation. It is the "root-mean-square deviation" originated by astronomers.

5. *Correlation* is the degree of agreement between two sets of data. In this section, the data will usually

be scores on two tests for the same individuals, or scores on one test and marks given to the same individuals by a teacher. Less often they will be correlations between scores on other measures—interest inventories, personality scales, self-concept ratings—and test scores or marks.

Correlation is expressed in terms of a *correlation coefficient,* generally designated by the symbol *r.* This is an abstract number that can take on values between 0 and 1.00. The value of 1.00, almost never found, shows perfect agreement in the rank order of scores on one variable and scores on a second variable. The value 0, as that figure implies, shows absence of relationship between two sets of scores or random association between the sets. When the coefficient is preceded by a plus sign (+) or is presented without a sign preceding it, the correlation is said to be positive, with high scores on the first variable being most often associated with high scores on the second variable and low scores on the two variables also being associated with each other. When the coefficient is preceded by a minus sign (—), the correlation is said to be negative. This occurs less frequently, as one might expect, for in such cases high scores on the first variable are associated with low scores on the second variable, and vice versa.

6. *Multiple correlation* is the degree of agreement between one variable, the criterion, and the best-weighted combination of a set of two or more other variables. An example would be the correlation between two test scores obtained at the beginning of a period of instruction—say, an achievement test score and an intelligence test score—and another test score at the end of instruction, generally an achievement test score in the same subject. A common example from outside the scope of this document would be the multiple correlation between high school average and entrance test scores used as predictors and grade point average at the end of the freshman year in college. Multiple correlation is expressed in terms of a *coefficient of multiple correlation,* designated by the symbol R to distinguish it from *r,* the symbol for simple correlation between two variables. This coefficient also takes on values between 0 and 1.00. When compared with the simple correlation between each of the predictor variables separately and the criterion, it shows the improvement in efficiency of prediction achieved by using the several variables in combination to predict the criterion. Multiple correlation R is always expressed without a sign because it can be used only to express the strength of a relationship.

7. A *regression equation* is an equation for predicting a criterion measure from the information provided by a single predictor or a set of two or more predictors. If a single predictor is used, we speak of simple regression or a simple regression equation; if two or more predictors are used, we speak of multiple regression, or a multiple regression equation. Correlation as described in definitions 5 and 6 preceding is the basis for determining the coefficients to be used in the equation.

CULTURAL BIAS IN TESTS

The concept of cultural bias is receiving new attention. In the late 1940's and early 1950's much professional effort was devoted to analyzing tests with a view to producing "culture-free" or "culture-fair" tests (Machover, 1943; Turnbull, 1949; Davis et al., 1951). Continuing efforts have been made by Cattell (1963) in his distinction between "crystallized" and "fluid" intelligence. Lorge (1952) pronounced a definitive evaluation of such efforts generally by pointing out that the major source of bias is to be found in society's "demands" and that tests must be related to those biases to define the cultural handicap of the disadvantaged in meeting the demands so that efforts may be directed toward correcting disadvantage and measuring progress in correcting it in individuals.

Two recent reviews, by Lambert (1964) and Anastasi (1964), merit mention as references here. Lambert summarizes information about a great variety of measures of aptitude and achievement designed to be "culture-fair" and includes much obtained from direct correspondence or conversation with interested researchers. Anastasi clarifies the relations among a number of the measures and, particularly, the concept of culture-fairness as that varies with different groups studied and purposes served. For example, she points out that:

> It is commonly assumed that nonverbal tests are more nearly culture-fair than are verbal tests. This assumption is obviously correct for persons who speak different languages. But for groups speaking a common language, whose cultures differ in other important respects, verbal tests may be less culturally loaded than tests of a predominantly spatial or perceptual nature.

Anastasi also points to factors that may normally be considered to limit the "culture-fairness" of a test, but have validity in a particular situation. Thus,

> ... the same factor that lowered the test score would also handicap the individual in his educational and vocational progress and in many other activities of daily life. Similarly, slow work habits, emotional insecurity, low achievement drive, lack of interest in abstract problems, and many other culturally linked conditions affecting test scores

are also likely to influence the relatively broad area of criterion behavior.

The reader should not be surprised, then, to find tests pronounced unbiased simply because they reflect the attributes that predict further achievement in school.

The view taken here separates society's demands into two chief parts: inescapable demands of living in an increasingly technological, urban, somewhat closed culture, and demands enforced by cultural distinctions of observable behavior largely associated with speech and historical knowledge. A current cigarette advertisement has capitalized on this by asking, "What do you want: good grammar or good taste?" A common speech fault in English is use of the double negative, a "fault" generally reenforced for the disadvantaged child by the constant pressure of his home and neighborhood; yet in most modern foreign languages, the double negative is correct usage to achieve emphasis. And American students have to learn to correct their fault of forgetting to use the double negative!

Spelling is another mark of cultural bias. Among the readers of a publication like this, or of any publication intended for general currency, unfavorable notice would certainly be taken by many of faulty spelling if at all frequent. Yet it is doubtful that the meaning would have been unclear, as witness the fact that others will read by each error without noticing it. It may be noted that spelling enjoys the status of a school subject only in English-speaking countries because English is the only language not uniformly phonetic. Early emphasis on formal approaches to correct spelling can intimidate an otherwise competent child from exercising a free flow of writing for fear of misspelling. How much better a situation in which a child writes to inform distant parents that he has an "earake," enabling the family to swing into action immediately. "What do you want: good spelling or good medicine?"

The effect of frequent correction for the "stigmata" of poor speech and poor spelling is subject to review and curricular revision if it is agreed that early overemphasis on correctness produces academic and affective deficiencies. Certainly, there is a distinction now being pondered between society's cultural demands that all be able to read, calculate, communicate, and acquire a background of structured knowledge in order to participate effectively in society, and society's cultural biases which have been illustrated here from grammar and spelling, but which go much deeper.

Having made the above observations to put the matter of cultural demands in perspective, it is necessary to return to the earlier observations attributed to Lorge and Anastasi. The tests themselves as of any date must be judged in terms of their validity for predicting the currently accepted goals under current procedures of instruction.

The discussion that follows of *Publishers' Test Information* is limited to a sample of tests that are representative of the sorts frequently used in ability grouping at various grade levels from preschool to college. Considerable detail is given about a few tests widely used in elementary and secondary schools in grouping and in evaluating achievement. In addition, the most popular measure for use at the preschool level, a major college entrance examination, and two new tests specially designed to meet the problems of testing minority children are discussed briefly. Thereafter the discussion proceeds to relevant research studies of less specific emphasis.

PUBLISHERS' TEST INFORMATION

The search for information about tests most widely used in school testing situations was initiated with a letter to each of seven major publishers of standardized tests asking for any data or other information they might have available about their own tests that would be pertinent to their use in ability grouping. Particular interest was expressed in predictive validity and/or reliability coefficients that the publishers themselves might have developed for groups differentiated by socioeconomic levels, or by race or ethnic background.

While only four of the seven publishers could provide useful data about tests on which they had done research, others reported research in progress, and all indicated that they were sensitive to the need for testing instruments free from cultural bias. Some reported the addition of members of minority groups to their professional staffs and provision for review of their test items by representative committees to detect instances of item bias.

Data supplied by test publishers are presented below. For some tests, only reliability data are available; for others, there are data regarding both reliability and predictive validity. With very few exceptions, these statistics show the tests to be unbiased with respect to any minority group, ethnic or socioeconomic; where such statistics favor one group over another, they appear to favor the minority rather than the majority group.

For the *Preschool Inventory,* formerly called the *Caldwell Preschool Inventory,* an instrument designed for use in the Head Start Program, Educational Testing Service reports deciles, summary statistics, and statistical characteristics for 317 children in eight kindergarten centers in North Carolina. This sample was divided into three groups by a consideration of each child's standing on two measures of socioeconomic status, the Coleman Index and an adaptation of the Ypsilanti *Cognitive Home Environment Scale,* itself an

Table 14
Clymer-Barrett Prereading Battery
Reliability Coefficients for Special Groups and Norms Group

TEST	SPECIAL GROUPS				NORMS GROUP
	B	C	D	E	
Visual Discrimination	.96	.97	.94	.97	.94
Auditory Discrimination	.94	.98	.89	.94	.82
Visual-Motor	.91	.94	.95	.95	.89
Total (Short Form)	.94	.97	.93	.96	.92
Total (Full Form)	.97	.98	.96	.98	.95

adaptation of Wolf's *Environmental Process Scale.* The two measures correlated .51 with each other. Scores for children at three socioeconomic status (SES) levels increased from the low to the high group but the differences in mean score were not significant. KR_{20}* reliability coefficients were .91, .89, .91, and .92 for low, middle, and high SES groups and the total group, respectively; for the total standardization sample, the KR_{20} reliability coefficient was .91. Individual items which appeared to be unusually difficult or unusually easy for the low SES group were, more often than not, the same items that were unusually difficult or unusually easy for the total North Carolina group and for the standardization sample.

In the Directions Manual for the *Clymer-Barrett Prereading Battery,* published by Personnel Press, Inc., split-half reliability coefficients are presented for four groups of first-grade children selected because of their difference from the norming population or because they might present special testing problems resulting in unreliable work on the tests. These groups are described as follows:

Group A Kindergarten pupils tested in May; 120 children in three classes, one system. Mean total score 74.85.

Group B First grades in three bilingual, rural schools in the Southwest; 63 pupils, mean total score 24.4.

Group C First grade in a rural, white, low-ability school; 52 pupils, mean total score 20.0.

Group D First grade in a rural, Negro, low-ability school; 28 pupils, mean total score 24.2.

Group E Five first grades in two mixed-ethnic, deprived neighborhood schools in a very large city: 111 pupils, mean total score 25.6.

*Kuder-Richardson reliability coefficients, Formula 20.

The reliability data for groups B, C, D, and E are presented above, together with those for the norms group. The data for Group A are omitted because they are for a group that is exceptional only in age (very young) rather than in cultural background.

The data indicate that even though the *Clymer-Barrett Prereading Battery* may be considerably more difficult for children in educationally atypical groups, it performs as well with them as it does with early first graders in the usual kinds of educational settings, so far as reliability is concerned.

By far the largest amount of data based on the use of tests with atypical groups has been published by Harcourt Brace Jovanovich, Inc. This is especially appropriate since their tests are used so widely in so many kinds of testing situations, especially those involving grouping.

For the *Metropolitan Readiness Tests,* the Manual of Directions provides split-half reliability data for seven different school systems at different socioeconomic levels with mean total scores ranging from 51 to 66. Since the subtests are so short that it is recommended that relatively little significance be attached to the subtest scores of *individual* students, only the reliability coefficients for total score are shown.

Alternate form, or test-retest, reliability data are also given for end-of-kindergarten children in systems D, E, F, and G. For both Form A first - Form B second and Form B first - Form A second groups, total score reliabilities of .91 are reported. With the observed reliability values for total score ranging from .90 to .95 and the measurement error of an individual score ranging from 3 to 5 points, as reported by the publisher, it would appear that total scores on the *Metropolitan Readiness Tests* may be used with considerable confidence for the purposes for which the tests are recommended.

Table 15
Metropolitan Readiness Tests
Split-Half Reliability Data for Form A in Seven School Systems

SCHOOL SYSTEM	NUMBER OF STUDENTS	GRADE	MONTH OF TESTING	MEAN SCORE	r_{11}
A	167	1	October	63.0	.91
B	173	1	October	57.9	.91
C	200	1	October	50.8	.94
D	88	Kdg.	May	66.4	.95
E	86	Kdg.	May	54.0	.93
F	59	Kdg.	May	53.4	.91
G	65	Kdg.	May	52.9	.90

Table 16
Metropolitan Readiness Tests
Split-Half Reliability Data for End-of-Kindergarten Administration of Form B in Systems D, E, F, and G

SCHOOL SYSTEM	NUMBER OF STUDENTS	MEAN SCORE	r_{11}
D	82	66.5	.93
E	91	53.2	.94
F	55	55.8	.92
G	61	51.0	.93

The manual also provides predictive validity data for a variety of student groups and circumstances. The basic data include correlations between readiness scores and scores on the *Stanford Achievement Test: Primary I* (1964 Revision) the following May for 9,497 students in the USOE First-Grade Reading Study of 1964-65 who participated in the standardization for the Readiness tests. Mitchell (1967) later used the scores of the same students to investigate the predictive validity of these tests and the *Murphy-Durrell Reading Readiness Analysis* by ethnic and socioeconomic differentiation. Certain of the Mitchell data, available upon request from the publisher, are summarized in Tables 17-19 on pages 64 and 65.

It is well to reiterate here the rationale of the statements above and below regarding bias in the tests. A test is adjudged to be biased only insofar as it provides information that leads to faulty inferences. If a test gives dependable evidence of present status on a variable for members of a minority group, as measured by a high reliability coefficient, and if it also predicts subsequent achievement as well for minority groups as for the general population represented in the norms, as measured by equally high correlation with achievement scores, the test is unbiased in its use for these purposes. The test may yield lower scores for minority group students, reflecting a disadvantage for the group on that test that is matched by the disadvantage these students experience in meeting the standard demands of instruction. Thus, the bias is in past conditions, or in the absence of effective adaptation of instruction, rather than in the tests.

The results shown in Table 17 do not support the hypothesis that the Metropolitan or the Murphy-Durrell tests have lower predictive validity for minority group students than for white students. For the Metropolitan tests, of the 15 correlations shown, 12 favor minority groups; for the Murphy-Durrell tests, nine of the 15 correlations favor the minority groups. Nor is there any consistent pattern of advantage or disadvantage among the three minority groups.

Table 17

Correlations between Total Score on Metropolitan Readiness Tests and Murphy-Durrell Reading Readiness Analysis, Administered to First Graders in Early October, and Scores on Reading Subtests of the Stanford Achievement Test the Following May, for Various Ethnic Sub-Groups of the Total Group of 9,497 Pupils Taking Both Tests

Correlations of Metropolitan Readiness Tests with Stanford Achievement Test: Primary I, Form X

Group	N	Word Reading	Paragraph Meaning	Vocabulary	Spelling	Word Study Skills	Standard Deviations of Metropolitan Readiness Scores
White	7,310	.58	.56	.59	.54	.59	15.8
Negro	518	.60	.55	.52	.56	.60	16.6
Mexican	139	.61	.56	.60	.57	.64	16.8
Oriental	37	.63	.51	.65	.60	.53	15.5
Ethnic origin unknown	1,473	.68	.69	.69	.66	.71	19.3
Total Group	9,497*	.63	.60	.63	.57	.64	17.5

Correlations of Murphy-Durrell Reading Readiness Analysis with Stanford Achievement Test: Primary I, Form X

Group	N	Word Reading	Paragraph Meaning	Vocabulary	Spelling	Word Study Skills	Standard Deviations of Murphy-Durrell Scores
White	7,310	.60	.58	.52	.57	.59	26.7
Negro	518	.63	.56	.52	.58	.61	25.5
Mexican	139	.58	.55	.59	.58	.61	26.1
Oriental	37	.68	.58	.62	.62	.50	24.4
Ethnic origin unknown	1,473	.69	.67	.63	.66	.69	29.9
Total Group	9,497*	.64	.61	.57	.60	.64	28.4

Standard Deviations of Stanford Raw Scores

Group	Word Reading	Paragraph Meaning	Vocabulary	Spelling	Word Study Skills
White	7.4	9.7	6.5	6.1	10.0
Negro	7.1	8.4	6.0	6.4	9.9
Mexican	7.1	8.6	5.8	6.3	9.8
Oriental	6.9	9.1	5.5	5.6	10.7
Ethnic origin unknown	8.7	10.7	7.3	6.9	11.8
Total Group	7.8	10.0	6.8	6.3	10.6

*The sum of the five N's above is only 9,477. The total group contained 15 Puerto Rican and 5 Indian-Eskimo children, for whom correlations were not computed.

In terms of socioeconomic differentiation, the predictive validities of the *Metropolitan Readiness Tests* appear to be considerably higher for the scores of children in less privileged communities than for those in more privileged communities. In comparing the predictive validities in Tables 17 and 18, however, it is important to consider the relative size of the standard deviations of the scores on the Readiness tests. The differences indicate greater variability for the readiness of children in the less privileged communities, and this would act to inflate the validities for these groups. Had the standard deviations for the two kinds of communities been more comparable, the differences in validities would have been less pronounced.

Table 18

Metropolitan Readiness Tests
Predictive Validities of Total Scores by Adult Level of Education
in the Child's Community

9 years or less, N = 1,411					13 years or more, N = 1,322	
Median Adult Level of Schooling in Community	**Stanford Achievement Test: Primary I, Form X**					**Standard Deviation, of Metropolitan Readiness Scores**
	Word Reading	**Paragraph Meaning**	**Vocabulary**	**Spelling**	**Word Study Skills**	
13 years or more	.57	.57	.59	.54	.57	14.4
9 years or less	.74	.70	.66	.64	.72	18.8

Standard Deviations of Stanford Raw Scores

	Word Reading	**Paragraph Meaning**	**Vocabulary**	**Spelling**	**Word Study Skills**
13 years or more	7.4	9.9	6.4	6.0	9.5
9 years or less	7.6	9.5	6.8	6.4	10.4

Table 19

Metropolitan Readiness Tests
Predictive Validities of Total Scores by Median Annual Income of Community

Above $8,000, N = 1,388					Below $4,000, N = 1,270	
Average Community Income	**Stanford Achievement Test: Primary I, Form X**					**Standard Deviation, of Metropolitan Readiness Scores**
	Word Reading	**Paragraph Meaning**	**Vocabulary**	**Spelling**	**Word Study Skills**	
Above $8,000	.65	.60	.60	.59	.626	14.5
Below $4,000	.71	.70	.68	.63	.712	19.1

Standard Deviations of Stanford Raw Scores

	Word Reading	**Paragraph Meaning**	**Vocabulary**	**Spelling**	**Word Study Skills**
Above $8,000	7.1	9.6	6.1	5.6	9.3
Below $4,000	7.7	9.6	6.7	6.5	10.6

For the *Otis-Lennon Mental Ability Test,* also published by Harcourt Brace Jovanovich, Inc., split-half reliability data are provided for five socioeconomic levels of community. These are shown in Table 20 on page 66.

In addition to the reliability data for different socioeconomic strata, the Technical Handbook accompanying the Otis-Lennon tests reports standard errors of measurement for successive score levels from IQ 50-70 to IQ 128-150. These range from 3.2 to 7.9 for single grades at single IQ ranges and from 4.4 to 6.6 for IQ level average, and average 4.9 for the total group.

Validity data for the Otis-Lennon test are reported for a large number of schools with mean IQ's as high as 110 and as low as 94. Correlations between Otis-Lennon scores and scores on several widely used achievement test batteries and ability tests and with end-of-year course grades are given. School districts tested are identified as to SES level. Correlations between Otis-Lennon scores and scores on the achievement tests range from .50 to .80; correlations between Otis-Lennon scores and teacher grades are somewhat lower; and correlations between Otis-Lennon scores and scores on other ability tests are somewhat higher.

Table 20
Otis-Lennon Mental Ability Test
Split-Half Reliability Coefficients for Socioeconomic Strata
of the National Standardization Sample

		Otis-Lennon Level and Grade					Number of
		Primary I Grade 1	Elementary I Grade 3	Elementary II Grade 5	Intermediate Grade 8	Advanced Grade 11	School Systems Within Stratum
Socioeconomic Level*							
High	Median	.87	.90	.94	.94	.94	9
	Range	.79-.90	.87-.95	.90-.95	.92-.95	.94-.96	
Above Average	Median	.88	.94	.95	.94	.94	11
	Range	.85-.91	.90-.95	.94-.96	.92-.96	.93-.96	
Average	Median	.90	.92	.94	.95	.95	17
	Range	.87-.93	.87-.93	.83-.96	.93-.96	.92-.97	
Below Average	Median	.91	.92	.95	.95	.94	9
	Range	.88-.93	.89-.94	.94-.97	.92-.97	.93-.96	
Low	Median	.90	.92	.95	.96	.95	8
	Range	.89-.93	.90-.94	.93-.97	.93-.96	.92-.96	
Complete Standardization Sample		.90	.92	.95	.95	.95	

*Public school systems with less than 300 total enrollment were not included in this analysis.

To aid in the interpretation of scores on the tests included in the College Entrance Examination Board Admissions Testing Program, the Board has published annually score report booklets for students, counselors, and admissions officers, and, periodically, much more comprehensive score reports. In addition, they have, through the years, commissioned a large number of research studies, and reports of many of these studies have found their way into professional journals. Two of these reports are particularly pertinent to the present discussion.

Studies conducted by Roberts (1962), Hills, Klock, and Lewis (1963), Boney (1966), and Stanley and Porter (1967) gave evidence that the *Scholastic Aptitude Test* (SAT) of the College Entrance Examination Board was as valid for predicting grades of students in predominantly black colleges as for predicting the college grades of white students (Kendrick and Thomas, 1970). The possible bias of the SAT in predicting college grades at integrated colleges was investigated by Cleary (1968) at the suggestion of the College Board.

Cleary and Hilton (1968) had earlier investigated possible bias in the *Preliminary Scholastic Aptitude Test* (PSAT) by studying the test items to see whether any items produced an uncommon discrepancy in scores for different racial and socioeconomic groups. On the basis of four separate studies of analysis of variance attributable to (1) "race," (2) SES, and (3) items, in the responses of 1,410 twelfth-grade students who had taken the PSAT in seven integrated high schools in three large metropolitan areas in 1961 (N = 636) or 1963 (N = 774), Cleary and Hilton concluded that while there were a few items producing an uncommon discrepancy between the performance of Negro and white students, the PSAT for practical purposes was not biased either for different ethnic groups or for groups at different socioeconomic levels. They based their conclusion on the absence of interaction* effects between item and "race" or item and SES.

*Interaction between two variables in an analysis of variance is a term to describe the tendency of individuals with particular combinations of status on the two variables to do much better or worse than would be indicated by their standing on the two variables separately. Here, if "race" or SES had given excessive disadvantage on particular items, the analysis of variance would have shown large interaction effects between item and "race" and/or item and SES.

The possible bias of the SAT in predicting college grades of black students at integrated colleges was investigated by Cleary (1968). She used the test as a whole as a predictor of college grade averages for both black and white students, hypothesizing that the test could be considered to be biased if too high or too low a criterion score was consistently predicted for members of the subgroup. Cleary concluded that there were no significant differences in prediction for black and white students from the two eastern colleges represented in the study. At a third college in the Southwest, significant differences were found in the regression lines for black and white students, but it was a matter of overprediction of college grades for black students by the use of the white or common regression lines.

In a study parallel to Cleary's, involving 13 integrated colleges, Temp (1971) found that the use of a regression equation based on the majority or white student group resulted in the prediction of college grades for black students that were higher than those that they actually earned. According to Temp, colleges might consider the possibility of using separate regression lines for black students.

As this document is being written, a comprehensive technical report on research and development activities relating to the tests in the College Board Admissions Testing Program is in press (William H. Angoff, ed.). In addition to an overview of administrative and technical problems of the program itself, the report describes construction practices involved in the SAT and the College Board Achievement Tests, discusses the statistical characteristics of the tests, the score scales, test validity, and the norms, and summarizes the results of several special studies having to do with the possible effect on test performance of coaching, test repetition, fatigue, anxiety, curriculum bias, and social and cultural factors. The Hilton and Cleary and the Cleary studies described above are among those reported.

A two-part Report of the Commission on Tests (College Entrance Examination Board, 1970) offers a variety of position papers, supported by research studies, on future directions for the College Board's program offerings. The commission of 21 members was drawn from persons variously concerned about and qualified to deal with emerging issues in the use and interpretation of the tests in that program. The papers in this compilation, covering a broad range of purposes and services, bear in varying degree on the issues under discussion here. In particular, the opening article of Part II. Briefs, by John Carroll, endorsed by 19 of the 21 commission members, recommends revision of the SAT to accomplish better descriptive measurement of college applicants, especially the disadvantaged. Hope is expressed that psychometric techniques might be applied to the development of tests that will provide for separate report scores for (1) verbal knowledge (culturally influenced), (2) reasoning ability (largely verbal but less influenced by breadth and richness of cultural experience), and (3) listening comprehension (a capability separately important and presumably less influenced by culture than reading), and (4) a de-emphasized section on quantitative reasoning (still hopefully allowing the culturally disadvantaged to show their potential as the present mathematics section does, relatively independent of verbal facility). The reader is directed to the original documents for the details which may be of particular interest and applicability in his own situation.

The American College Testing Program (ACTP), which seeks to serve the same function in college admissions, has its own intensive research studies in progress designed to identify item and/or test bias in its offerings. A major technical report, incorporating the findings of these studies, will likewise seek to map a course for the ACTP but is not scheduled for publication until late 1971 or early 1972.

Two new tests designed especially for use with the disadvantaged have recently been reported in the literature: a *Reading Prognosis Test,* published by the Institute of Developmental Studies, and the *Orr-Graham Listening Test,* also known as BoLT for Boys' Listening Test, published by the American Institutes of Research.

The *Reading Prognosis Test* is a 25-minute test, individually administered, measuring language, perceptual discrimination, and beginning reading skills. In a series of studies, the test was pretested and validated on balanced samples that included equal numbers of children from middle and lower socioeconomic groups and equal numbers of Negro and white children (Weiner and Feldmann, 1963). In an initial pilot study involving 40 children, the *Reading Prognosis Test* correlated .87 with the *Gates Primary Reading Tests: Word Recognition* of 1958. A second study involved 126 children, tested in October with the new test and in May with the *Gates Primary Reading Tests: Sentence Reading* and *Paragraph Reading.* In the October testing, retesting within three weeks of the initial testing yielded a reliability coefficient of .93 for the total group. At this time also the concurrent correlation with the *Lorge-Thorndike Intelligence Tests* for 138 children was .42 for the lower SES group and .21 for the middle SES group. The correlations of the Reading Prognosis total test score with the Paragraph Reading test ranged from .79 for the lower-class Negro female group to .89 for the middle-class white male group. The total group correlation was .81. The correlations of the Reading Prognosis total test score

with the Sentence Reading test ranged from .61 for the middle-class Negro female group to .88 for the middle-class white female group. The authors concluded that the Reading Prognosis total test score, at the beginning of Grade 1, is a good predictor of Gates scores for different SES groups at the end of a year's instruction.

In a later validation study involving 300 Negro and white first graders in a large urban area and in a suburban community, correlations between the *Reading Prognosis Test* and the *Gates Primary Reading Tests: Paragraph Reading* and the *Metropolitan Reading Test* at the end of Grade 1 ranged from .71 to .80, and correlations for separate ethnic and SES groups from .66 to .88 (Feldmann, 1965). Other and largely similar validation data are reported in the 1964-65 Research Memos of the Institute of Developmental Studies. Generally, the best prediction is shown to be for Negroes and for the lowest SES group.

The *Orr-Graham Listening Test* was developed between 1964 and 1968, with the financial support of the College Entrance Examination Board, to identify educational potential among disadvantaged eighth-grade Negro boys. The content of the test, an 86-item, 90-minute instrument, administered orally, was designed to be of interest to boys of junior high school age. The stories in the test are based on such topics as spies, baseball players, cowboys, and soldiers. The test was developed to elicit motivation through increased interest and to provide a test of aptitude which was not dependent upon reading proficiency.

All research, from that preceding the actual development of the test, through preliminary tryouts to the final administration, was carried on in junior high schools in the District of Columbia. About 99 percent of the boys included in the samples were Negroes. On the basis of a "final administration" of the test, Orr and Graham (1968) reported the test to be reliable, acceptable to the group for which it was intended, and uniquely different from the traditional aptitude and achievement tests. They obtained a split-half reliability coefficient of .85 and a KR_{20} reliability coefficient of .89. Correlations of the total test score with total scores on the *School and College Ability Test* (SCAT), *STEP Listening,* and *STEP Reading* were .60, .49, and .69, respectively. The results showed that about 81 percent of the boys liked the Listening test and preferred it to a reading test covering the same content.

Carver (1969) reported on a replication of the Orr and Graham study with extension to other ethnic and income-level groups. In this study, 615 eighth-grade boys in the District of Columbia area, 314 Negroes (182 low-income, 132 middle-income) and 301 whites (110 low-income, 191 middle-income) were administered the Listening test, SCAT (Level 2), and *STEP Listening,* and filled out questionnaires. Family incomes of $5,000 divided the low- and middle-income groups.

An incidental reliability study of 142 low-income Negroes yielded an alternate form reliability of .78. For the low-income Negro group, correlations between the Listening test and other test variables were highly similar to those in the earlier study; for all groups combined, the Listening test correlated .69 with SCAT total score and .78 with *STEP Listening,* considerably higher than the correlations in the earlier study. The correlations between the Listening test and *STEP Listening* ranged between .65 for the low-income Negroes and .79 for the middle-income Negroes. The low-income Negroes scored lowest on all tests, the middle-income whites scored highest on all tests, and the difference between these two groups was always greater than one standard deviation. The questionnaire responses showed that all four groups preferred the Listening test to SCAT, but only the two Negro groups preferred it to *STEP Listening.*

Carver concludes that the reliability of the *Orr-Graham Listening Test* for low-income Negroes appears to be adequate and stable since there is little difference in the split-half correlations of the earlier study and the alternate forms correlations in his study. The concurrent validity is quite high, as indicated by the high correlation between the test and *STEP Listening.* The test also appears to be an adequate indicator of aptitude since the correlation with SCAT is high. He questions the high uniqueness of the test for identifying educational potential among the disadvantaged; to Carver the test is unique only in that it is preferred by Negroes. He finds no support for the hypothesis from the earlier test results that the effect of disadvantagement may be more associated with reading proficiency than with verbal proficiency in general. The large Negro-white differences are apparent in the Listening test as well as in the reading and verbal measures.

In two other articles (1968, 1968-69) Carver further discusses the questionalbe uniqueness of the test and the failure of the test to lessen score differences between Negroes and whites.

To summarize, systematic efforts are being made by test publishers and research agencies to review present test offerings and to introduce new emphases to meet the problem of assessing the capabilities of disadvantaged children. To date, the studies of old and new materials suggest possibilities but little accumulated capability for meeting the assessment problem directly.

The negative evidence that tests standardized on other populations tend to overpredict the subsequent performance of disadvantaged individuals, hence are not unfair to them, is cold comfort. The challenge is

to mount a campaign of innovative teaching and evaluative research that will enhance learning by describing learning progress directly, rather than to settle for procedures that are fair only in the sense that they reflect "fairly" the current unmitigated disadvantages.

Now that the problem of assessing the potentiality and achievement of variously disadvantaged children is being faced, we must trust to continuing honest effort to separate the essential from the secondary objectives of public instruction to provide differential criteria of effectiveness of instructional adaptations. Thereby, it should be possible to help those operating from limited backgrounds to achieve increasingly greater mastery of essentials, including a self-respect that allows them to make a distinction between the essential and the ornamental outcomes of education.

RESEARCH REPORTS ON THE USE OF TESTS WITH THE DISADVANTAGED

A second source of information, and a valuable one, was the Information Retrieval Center for the Disadvantaged at Teachers College, Columbia University. Useful studies found there were concerned with the testing of the culturally limited at all levels, from preschool to college students and adults; the testing of non-whites, including the Negro, the Mexican-American, and the American Indian; and the advantages and disadvantages of particular tests and particular types of tests for use with non-middle-class white groups.

Public libraries and university libraries gave access to the many periodicals in which articles were located through the *Education Index,* and to *Dissertation Abstracts* and *Psychological Abstracts.* The libraries of two test publishers proved a good source for unpublished studies. A visit to the Institute for Developmental Studies resulted in the location of other pertinent data, ERIC abstracts for reports related to *disadvantaged* and *testing* were examined.

Research relating to the effects of cultural background on test scores and the kinds of educational opportunities that have been afforded or denied the disadvantaged as a result of test performance has increased in volume and intensity as concern for the improvement and extension of opportunities generally for minority groups has become universal. But research of this kind is not new; for more than 60 years, researchers have been exploring and reporting the complexities and problems of the use of tests with culturally different groups, even though for much of that time what they had to report may have been listened to by relatively few. While the great bulk of this research has been reviewed in preparation for the writing of this document, no attempt has been made to summarize the research that has been summarized elsewhere, except for those studies that have particular pertinence here. Instead, emphasis has been put on those studies which have been done since 1960, most of them since 1965. Anyone interested in wider reading, particularly of the earlier studies, is referred to a half dozen of the most comprehensive surveys of the literature.

Lucas (1953) reviewed 253 pieces of literature relating to the effects of cultural background on scores on aptitude tests. Campbell (1964) included 46 references in his review of research done between 1932 and 1963 concerning the testing of culturally different groups. Pettigrew (1964) in the bibliography in his book on the Negro American listed among his 565 references almost 200 studies related to Negro-American intelligence. Shuey (1966) reviewed 382 studies in the latest edition of her volume bearing on racial differences in intelligence; while her conclusions relative to differences between Negroes and whites, as determined by intelligence tests, have been the subject of considerable criticism, few would contest the statement that her coverage of the literature of the last 50 years is extensive. Dreger and Miller (1968) reported a comprehensive survey of psychological studies of Negroes and whites done in the United States between 1959 and 1965. Flaugher (1970), in a recently completed review of research on testing practices, minority groups, and higher education, lists 65 references covering the years 1913 to 1970.

Studies of discrimination against minority groups in testing have usually dealt with the aspects of test content, the norms population, and the interpretation of results. What about the testing procedure itself? Do certain testing conditions systematically favor one cultural or racial group over another—examiner's race, test directions, pretest practice, speededness, test-wiseness? The next five studies were concerned with some of these conditions.

Pelosi (1968) made a study of the effects of examiner race, sex, and style on the test responses of adult Negro examinees. In his experiment, 96 Negro males were given six subtests of the *Wechsler Adult Intelligence Scale* (WAIS), the *Purdue Pegboard,* and the *IPAT Culture Fair Intelligence Test,* eight tests involving 12 scores, by examiners who included Negroes and whites, males and females, "warm" and "cold" personalities, with three examiners within each race-sex category. A separate analysis of variance was done for each of the 12 scores. None of the examiner attributes or the interactions between them were significant on seven of the eight tests. The exception was the Culture Fair test, group administered, for which "cold treatment by male Negro examiners resulted in substantially higher scores than those obtained by female Negro examiners." On all but one subtest of

WAIS, the mean scores were higher with white examiners and for examinees treated coldly.

Pelosi writes: "Though differences were small and non-significant, the general direction contradicts the findings of previous research which suggested inadvertent negative bias due to white examiners." He suggests two weaknesses in the study, however: (1) The subjects were volunteers, enrollees in an anti-poverty work experience project, and were not as "ego-involved" as would be the case in an actual testing situation. (2) The "warm" and "cold" examiners were not sufficiently different in the testing situations.

Abramson (1969) examined the effect of the race of both children and examiners on the child's performance on the *Peabody Picture Vocabulary Test,* an individually administered test. Two white and two Negro examiners administered the test to 88 and 113 white and Negro children in first grade and kindergarten, respectively, in an integrated urban school. The first graders had been in the school since their kindergarten year and the kindergartners had been in school for five months. The children had usually seen the examiner, a paraprofessional working in the school, at least once a day during the time they had been in school. The investigator found a small but statistically significant interaction of the examiner's race and the child's race for first graders but not for kindergartners. He suggested that this difference might have been the result of the first graders having reached an age of racial awareness, but there were no data available regarding racial awareness.

A study reported by Dublin and Osburn (1969) was directed toward investigating whether or not two other conditions, aspects of the test procedure itself—extra preliminary practice and extra testing time—systematically favored white examinees over Negro examinees. Their sample included 235 Negro and 232 white students, representing both high and low socioeconomic levels, from two high schools in Galena Park, Texas. All students in the sample were quite familiar with standardized tests. The *Employee Aptitude Survey* (four subtests) was used. Groups within each race in grades 9 and 10 were given the test with regular time limits; in grades 11 and 12 extra time was allowed. Some groups took only one form of the test; other groups took both forms, with the first testing considered as practice. An analysis of variance was done.

The order of mean scores was as follows:

BY SES AND RACE	BY TESTING CONDITIONS
High SES Whites	Power test with practice
Low SES Whites	Power test without practice
High SES Negroes	Speeded test with practice
Low SES Negroes	Speeded test without practice

Interesting findings of the analysis of variance were these:

1. Extra practice was no more advantageous to Negro than to white groups.
2. Both SES groups profited from extra practice to a comparable degree.
3. When Negro and white groups, matched by sex, grade level, and SES were compared, *improvement* in score from speeded to power tests was no larger for Negroes than for whites.
4. High and low SES groups profited equally by the tripled time limits.
5. When both extra practice and extra testing time were given, again the *improvement* was not significantly related to either race or socioeconomic status.

The authors concluded that the results implied in a general sense that "testing procedure itself is not a major factor in discriminating between culturally advantaged and culturally disadvantaged students."

Lo Monaco (1969) studied four groups of disadvantaged ninth-grade Negro boys to determine their response levels to both standard and oral-visual administrations of two vocationally relevant instruments. The boys were assigned to two experimental and two control groups equated for age, reading comprehension, and socioeconomic level. Hypothesizing that reading deficits contaminate scores on standard versions of the instruments and that disadvantaged youth have better listening comprehension abilities than reading ability, Lo Monaco administered three measures—the *Metropolitan Reading Test* (MRT), the *Kuder Preference Record-Vocational,* and the *Life-Planning Questionnaire-Modified* (LPQ-M)—to all groups in the standard version and in a modified oral-visual version involving no reading. The two experimental groups took both the standard version and the oral-visual version in difference sequence; one control group took the standard version twice, and the other the oral-aural version twice.

Except for the Reading test, oral-visual version scores were higher than the standard version scores on all measures; on the MRT, this was true for the low reading cases only. The oral-aural version provided more reliable measures of interests on the Kuder and of strivings on the LPQ-M than did the standard version. According to Lo Monaco, "the findings of this study indicate that reading deficits are important response variables...." Instruments can be modified to "mediate these difficulties."

Buchanan (1969) studied the effect of cultural deprivation on the approach to test-taking as indicated by response style to multiple-choice questions. Buchanan asked whether his social background, deficient education, and experience of failure would lead the

deprived student to reject the problem-solving approach when he is faced with questions to which he does not know the answers; that is, does he guess indiscriminately rather than attempt to eliminate the less plausible distractors in multiple-choice questions to arrive at an "educated" guess, as non-deprived students do?

Buchanan used three different tests at one grade level and one test at three different grade levels and analyzed (1) items on which non-deprived and deprived students experienced equal difficulty and (2) items with matched difficulty indices. For matched questions there was no difference between sub-cultural groups in the degree of selective guessing. Buchanan concluded that indiscriminate guessing is related to a real informational deficiency rather than to differences in motivation.

In a case study of the effects of educational deprivation on southern rural Negro children, Green and Hoffman (1965) worked in the public schools of Prince Edward County, which were closed from 1959 to 1963. During these four years, most Negroes had no schooling (No Educ group); some had an average of one and one-half years (Educ group). After resumption of school operation, the *Stanford-Binet Intelligence Scale* and the *Stanford Achievement Test-Partial Battery* were given to 154 children in the No Educ group and 125 children in the Educ group. Extensive tables given by chronological age in the Green and Hoffman report show that the extended educational deprivation had a depressing effect upon achievement and intelligence at all ages. Language deficits on the Stanford-Partial were greater than in other areas. On the Stanford-Binet, the differences between IQ's of children at the earlier ages who had had no schooling and those who had had some schooling were as great as 30 points. In both the No Educ and the Educ groups, there was a negative relation between age and measured IQ.

Goldstein et al. (1970) studied the effect of a specially designed, enriched curriculum for 161 children on (1) average test performance over the two-year range from beginning pre-Kindergarten to end of Kindergarten, and on (2) stability coefficients over the same range for Stanford-Binet IQ, the *Peabody Picture Vocabulary Test,* and the *Columbia Mental Maturity Scale.* Treating these three measures as measures of various aspects of cognitive development, they concluded that although mean gains on all three measures were reliable, the PPVT was not sensitive to effects of special instruction of these young disadvantaged children.

Lesser et al. (1965) studied the influences of different social classes and cultures on patterns among mental abilities: verbal, number, reasoning, spatial. They tested 320 first-grade children, including middle- and lower-class Chinese, Jews, Negroes, and Puerto Ricans, in New York City and New Rochelle, New York, with the *Hunter Aptitude Scales,* designed for gifted four- and five-year-olds. Social class was based on the *Hollingshead and Redlich Index,* using occupation, residence, and education of the head of the family as criteria. The scales were administered individually by well-trained psychometricians of the same ethnic group as the child.

Split-half reliabilities for the different ethnic groups ($N = 80$ for each group) ranged from a low .80 for Jewish children on Space to a high .96 for both Negroes and Puerto Ricans on Numbers. Split-half reliabilities by social class ($N = 160$ for each class) ranged from a low .80 for the middle class on Space to a high .96 for the lower class on Numbers. The middle-class children were slightly higher on Verbal but lower on Reasoning, Number, and Space. No tests for significance across ethnic or social-class differences were reported.

Means by ethnic group and social class are given in Table 21 below. The greatest differences in standard deviation were in Verbal.

An analysis of variance was done, and interactions of social class, ethnic group, and sex reported. The major findings were that (1) differences in social class *do* produce significant differences in absolute level of *each* ability, but *do not* produce differences in the pattern of abilities; (2) differences in ethnic-group membership produce differences in both absolute level and pattern of abilities; (3) social class and ethnicity interact to affect the level of each ability, but do not interact to affect patterns. The authors concluded by proposing that "the identification of

Table 21
Hunter Aptitude Scales

	Means for Ethnic Groups				Means for Social Classes	
	Chinese	Jews	Negroes	Puerto Ricans	Middle Class	Lower Class
Verbal	71.1	90.3	74.3	61.9	76.8	65.3
Reasoning	25.9	25.2	20.4	18.9	27.7	24.2
Number	27.8	28.5	18.4	19.1	29.8	25.6
Space	42.5	42.5	34.4	35.1	44.9	40.1

relative intellectual strengths and weaknesses of members of different cultural groups become a basic and vital prerequisite to making enlightened decisions about education in urban areas."

Brazziel and Terrell (1962) conducted an experiment in the development of readiness in a culturally disadvantaged group of first-grade Negro children, most of them from sharecropper homes. Twenty-six of the children were assigned to an experimental group and the other 66 to three control groups. Parents of the children in the experimental group were involved in registration and in the development of readiness activities. The experimental group was given a six-week readiness program, which involved travelogues, 30 minutes of educational television each day, and intensified activity to develop preception, vocabulary, and the will to follow directions. Weekly tests were given on some form of readiness.

At the end of six weeks, the *Metropolitan Readiness Tests* were given to both experimental and control groups. The test results of the experimental group were greatly superior to those of the control group, the percentile rank for total score for the experimental group being 50 as opposed to 16, 14, and 13 for control groups A, B, and C, respectively. The mean IQ of the experimental group in the spring of Grade 1 was 106.5, while second-grade Negro children in the country averaged 91.4 in the state testing program. Brazziel and Terrell attributed the success of the program to "an efficacious combination of direct teacher-parent partnership, excellent materials, test wisdom development, and energetic, uninhibited teaching. . . ."

Dowd (1968) studied sex and race differences in the effectiveness of various composite predictors of initial reading success. He tested 366 children from a large suburban district at the end of Kindergarten with the *Metropolitan Readiness Tests* (MRT), both the 1949 edition and the 1965 Revision, the Clark and Ozehosky *U-Scale* measuring self-concept, and the *Van Alstyne Picture Vocabulary Test.* At the end of Grade 1, he gave the *Gates Primary Reading Tests: Word Recognition* to 232 of the original 366 children still in school. For all groups (Negro, white—boys, girls) the best predictor was the MRT, except for the 1965 Revision for Negro boys; for them a combination of the Numbers and Copying subtests in the 1949 edition of the MRT provided the best prediction for the Gates tests. The *U-Scale* added significantly to the prediction in some instances; the Van Alstyne test did not.

Beidler (1968) worked with 276 students in Kindergarten through Grade 2 in two schools in a disadvantaged neighborhood in Bethlehem, Pennsylvania, to determine the effects of the use of the *Peabody Language Development Kits* (PLDK) on the primary grades. The experimental groups had seven months of use of the kits in addition to the normal language arts program followed by the control groups. The *Lee-Clark Reading Readiness Test* was administered to the Kindergarten in the spring, and the *Otis-Lennon Mental Ability Test* and the *Cooperative Primary Tests* in Reading and Listening to grades 1 and 2. A writing sample, scored for quantity and maturity, was obtained from grades 1 and 2.

At the kindergarten level, there was a highly significant differences in favor of the control group, leading one to suspect that the experimental and control groups at that level may not have been initially comparable. For grades 1 and 2, no significant differences were found on intelligence, reading, or listening scores; in Grade 2, however, the experimental group "wrote a significantly greater number of running words than did the control group." Beidler described the implications thus: ". . . compared to conventional procedures, seven months of PLDK lessons do not significantly improve the intelligence, reading, listening, or writing of disadvantaged children in the primary grades."

Harris and Lovinger (1968) reported somewhat different results from a longitudinal study involving 35 boys and 45 girls in a very disadvantaged area in the borough of Queens, New York City, in a school which had the lowest achievement and highest transiency rate of any junior high school in the borough. All 80 students had been given the same tests from the first grade on: Grade 1, *Pintner-Cunningham Primary Test;* Grade 3, *Otis Quick-Scoring Mental Ability Test: Alpha Level;* Grade 6, *Otis Quick-Scoring Mental Ability Test: Beta Level;* Grade 7, the *Wechsler Intelligence Scale for Children* (WISC); Grade 8, the *Cattell Culture Fair Intelligence Test* and the *Pintner General Ability Test;* Grade 9, WISC. There were 12 measures in all.

No decrease in IQ was found throughout successive grades for this group of disadvantaged Negro adolescents. Mean IQ at Grade 1 was 98, then 94, 88, 93, 96, 92, to 96 at Grade 9. On the WISC this group was not any more handicapped on verbal than on nonverbal tests. At Grade 7 the mean was 93.8 for Verbal and 93.7 for Performance; at Grade 9 the means were 96.1 and 97.0, respectively. The correlations between the tests given two years apart were .87 for Verbal, .85 for Performance, and .89 for Full Scale.

In 1962 a study of socioeconomic status and school achievement was made by the California Elementary School Administrators Association. The *School and College Ability Test* (SCAT) and the *Sequential Tests of Educational Progress* (STEP) were given concurrently to 3,008 sixth-grade students in 40 schools in three school districts. Grouping in terms of socioeconomic level was accomplished by use of the *Hollingshead*

Two-Factor Index, based on parent occupation and education level. The two top groups, of five, were combined to make four SES levels.

Of pertinence here are the correlations between SCAT and STEP by SES levels. Was the prediction equally good at all levels? The correlations between SCAT-Verbal, SCAT-Quantitative, and SCAT-Total and six STEP subtests by SES levels all followed the same general pattern. For all 18 sets of correlations, the lowest *r*'s were for the highest SES level. For 11 sets of correlations the highest *r*'s were for the next to the lowest SES level. For none of the 18 sets of correlations were the *r*'s for the lowest SES level as low as those for the highest SES level. In other words, the prediction was generally better for the lower SES levels than for the higher SES levels. The correlations between SCAT-Total and STEP by SES levels, from high to low, are given below.

Roberts et al. (1965) investigated the commonly reported tendency of Negro IQ's to drop with increasing age in a longitudinal study of the performance of 69 Negro-American children on the *Stanford-Binet Intelligence Scale,* with special concern for the "cause or associated factors" of the observed differences. In this study different forms of the test were administered to the children at age 5 and age 10, with the second examiner having no knowledge of the earlier results. Data were gathered on parent occupation, family pattern, and socioeconomic level.

Over the five-year period, male mean IQ's fell from 96 to 88 and female mean IQ's from 94 to 84, with the decreases being statistically different in both cases. The respective standard deviations were 17.5 and 21.4 for the males, a large increase, and 13.2 and 15.4 for the females. The decline in IQ for boys seemed to be related to low socioeconomic status and unstable and unfavorable family patterns; the decline in IQ for girls was slightly in reverse. The number of cases, however, was so small for the subgroups that little confidence can be placed in the statistics reported. The largest decreases were with children showing the greatest difficulty with verbal skills. Verbal Absurdities was an "outstanding failure." There was slightly less difficulty with Repeating Digits, and Making Change was relatively easy. None of the children tested at age 10 could pass the 10-year vocabulary test.

To obtain normative data on intelligence and achievement for a large homogeneous sample for which there were no previous data, Kennedy et al. (1963) administered the *Stanford-Binet Intelligence Scale* and the *California Achievement Tests* (CAT) to a well-selected sample of 1,800 Negro students in grades 1 through 6 in five southeastern states. They reported results by metropolitan, urban, and rural counties, age, sex, grade level, and socioeconomic status. For the entire sample the mean IQ was 80.7, with a standard deviation of 12.4. The mean IQ decreased with age, with type of community (from metropolitan to rural), and with socioeconomic level (from high to low); it remained relatively stable by grade. The order of the items by difficulty was quite similar to that of the norming population. The Negro students were relatively high on Rote Memory, Digits, Making Change, and Days of the Week, and low on Abstract Verbal, Vocabulary, Absurdities, and Comprehension. On the CAT the mean grade equivalent on the total battery fell increasingly below the norm (from .2 in Grade 1 to 1.2 in Grade 5) and decreased with socioeconomic level; there was, however, no difference in achievement by type of community. The correlation of the total battery with the Stanford-Binet mental age was .69, about the level usually found for total school groups.

Hughes and Lessler (1965) compared the *Wechsler Intelligence Scale for Children* (WISC) and *Peabody Picture Vocabulary Test* (PPVT) scores of 137 Negro and white rural school children of the lowest socioeconomic level in North Carolina. Ranging in age from 6 to 16, these children had been sent for testing because of suspected mental retardation. Could the shorter PPVT be substituted for the WISC, usually given? Correlations between the two tests ranged from

Table 22
California Correlations between SCAT-Total and STEP
by SES Level

SCAT-Total

		N	Mathematics	Science	Social Studies	Reading	Listening	Writing	Standard Deviation SCAT
STEP									
SES	A	524	.71	.62	.67	.64	.57	.61	10.7
	B	566	.78	.72	.75	.72	.66	.70	11.3
	C	524	.81	.78	.80	.76	.67	.74	9.0
	D	553	.76	.74	.79	.77	.66	.69	7.6

a low .21 for white males for PPVT with WISC Performance to a high of .66 for Negro males for PPVT with the Full WISC. Seven of the 12 correlations were .55 or higher. All but one of the *r*'s was significant at the one percent level and that one was significant at the 5 percent level. Generally, the *r*'s for Negro children were higher than for white children. With the standard error of estimate* running from 7 to 14 points, the authors concluded that "the PPVT has a distinct advantage over group tests of intelligence for these rural children... and would perform an adequate screening function when used in the school or by personnel from the mental health clinic." Assign the children, particularly disadvantaged rural children, to Educable Mentally Retarded classes on the basis of a vocabulary test!

An investigation by Kneif and Stroud (1959) was planned, first, to provide data on the social class or culture bias in intellectual testing and, second, to ascertain interrelationships among certain relatively new intelligence tests and tests of scholastic achievement. The *Lorge-Thorndike Intelligence Tests* (L-T), Verbal and Nonverbal, the *Davis-Eells Games,* Raven's *Progressive Matrices* (RPM), and the Warner *Index of Status Characteristics.* All tests except the RPM were administered to a sample of 344 fourth-grade students in a midwestern city, all the students present at the time in six of 18 elementary schools. One hundred sixty-four of these students who were in the fifth grade the following year were given the RPM.

All of the intelligence tests and composite scores on the *Iowa Tests of Basic Skills* (ITBS) correlated significantly with social status and, with the exception of the RPM, to approximately the same extent. The L-T Verbal scores gave the best prediction of ITBS scores, followed in order by L-T Nonverbal scores and the *Davis-Eells Games.* The L-T Verbal scores alone correlated with ITBS about as well as did the entire battery of tests when combined in multiple-correlation design. The RPM correlated to a smaller degree with ITBS than did any other intelligence test. The analysis gave little justification for the use of L-T Nonverbal, the *Davis-Eells Games,* and RPM in conjunction with L-T Verbal for general prediction purposes. This is not to deny, however, their usefulness in individual diagnosis.

Davis (1969) followed 103 randomly selected students from Grade 3 through grades 5 and 6 to "measure improvement in test performance in disadvantaged inner-city poverty tracts" in Knoxville during a federally sponsored Communication Skills Project. The *Metropolitan Achievement Tests* (MAT) of Reading, Word Discrimination, Language Usage, and Spelling were administered in Grade 3 in 1965. Improvement was measured by relating to the 1965 results 1966 and 1967 scores from *California Achievement Tests* (CAT) in Reading Vocabulary, Reading Comprehension, Mechanics of English, and Spelling. Davis reports that "over the three test periods 48 comparisons for significance of differences... were run. Computed results indicated significant differences in thirty-two of the forth-eight comparisons."

Davis states in his thesis that a basis for comparability of the MAT and CAT subtests were accepted when given correlation coefficients between areas of the two tests ranged from .77 to .95. It should be pointed out that correlation indicates only similarity in rank; it tells nothing of the grade equivalent scores, which could differ by months for students taking the two tests. There are also questions as to how standard scores and raw scores could be compared across the two tests (and levels) as the Grade 3 results on the MAT were compared with Grade 4 and Grade 5 results on the CAT. Was "improvement" the gain from Grade 3 to later grades in the achievement areas considered? This comparison of results across different tests is very common even though not proper. There is evidence that MAT and CAT, particularly, are not comparable as to grade equivalent scores. CAT gives higher results and grade equivalent scores have a much smaller standard deviation.

The report appears to be an attempted evaluation of the effect of a federal project. How could this be measured by using gain over two years? There appears to be no relation of the gains to those of a group not in the study. What gains over the same period of time for the same schools have been made in previous years? What national norms give 1.0 as a normal yearly gain?

A study of Eagle and Harris (1969) examined the relationship between race and performance on two standardized reading tests, the reading tests of the *Iowa Test of Basic Skills* and the *Metropolitan Achievement Tests.* The tests were administered to 850 fourth-grade students and 850 sixth-grade students in all elementary schools of an urban district near New York

*The standard error of estimate is simply the standard deviation of the differences between scores of the same individuals on the criterion test and the predictor test, in this case expressed as IQ's. It is to be distinguished from the standard error of measurement, which accepts the test being studied as its own proper criterion and seeks to estimate departure of the value found on this test from the hypothetical true value that this test measures imperfectly because it cannot be made infinitely long. See definition on the standard error of measurement on page 59.

City. Although white students earned higher scores than nonwhite students on both tests, the Metropolitan produced significantly greater differences between the races, at both grade levels, than did the Iowa. At Grade 4, the Metropolitan gave white students a superiority over nonwhite students of .72 compared to .58 years for the Iowa. At Grade 6, however, the Metropolitan gave white students a superiority over nonwhite students of 1.13 years compared to .73 for the Iowa, a difference of about five months. Analysis of variance confirmed the statistical significance of these differences at both grade levels.

In brief, the Eagle-Harris findings imply that white elementary school children are "favored" by the Metropolitan whereas Negro children are "favored" by the Iowa when results are contrasted. Why is this so? Must one question the validity of one or the other of these highly respected tests? The authors suggest that in previous investigations involving comparisons among standardized achievements tests, little consideration has been given to the question of interaction effects between tests and sociocultural variables. Yet, failure to take into account significant interactions can mask important changes taking place in subgroup student performance and could provide the basis for erroneous or misleading evaluation of curriculum effectiveness.

The implications of findings like those of Eagle and Harris could be profound. With the knowledge that one test would be more reflective of gains for a particular subgroup than another, what administrator would not choose to use the test that demonstrates the kind of performance, maximal of minimal, that will best suit his practical purposes?

Santos (1967) studied the level and variability of achievement in educationally disadvantaged attendance centers in Iowa, and investigated item characteristics of the *Iowa Tests of Basic Skills* (ITBS) between educationally disadvantaged and total representative groups. In the Iowa 1966 testing program with ITBS, the educationally disadvantaged schools in all grades and all test areas were almost a year below the norm for representative schools. Difference in item difficulty between representative and disadvantaged schools was pronounced, and quite variable. The discrimination indices were equally satisfactory in the two groups. Santos suggests that research with experimental programs implies a need for reducing cultural bias, adapting content to needs and interests, and adjusting the difficulty of the test materials. "At the present time statements of behavioral objectives... are not specific enough to be of much help to authors of achievement tests in determining content, emphasis, and grade placement."

Alzobaie et al. (1968) administered the *Lorge-Thorndike Intelligence Tests,* Verbal and Nonverbal, three of Guilford's tests of creativity, the *Test of Academic Performance-Reading,* and two scales from the *Cattell Culture Fair Intelligence Test* to 122 disadvantaged tenth-grade Negro students, in a district adjacent to Watts in Los Angeles. Grade point averages (GPA) and SES indices from the Warner *Index of Social Class* scale were also obtained for each student. Intercorrelations among the predictors ranged from .23 to .82; the Guilford total score had correlations ranging from .40 to .56 with the other predictors. The Lorge-Thorndike and Reading tests showed small but significant correlations with SES; the Guilford and Cattell tests did not. Correlations with a convergent criterion measure* of academic success, GPA, ranged from .29 and .32 for the Cattell scales to .56 for the Reading test; correlations with CPA for the three Guilford tests, essentially divergent tests, were .46, .39, and .31, with .48 for the composite. The authors concluded:

> Despite their brevity, the three essentially nonverbal tests of divergent production as well as their composite score showed promise in the prediction of GPA. Thus, the three Guilford tests afford an alternative means for predicting traditionally evaluated academic performance of culturally disadvantaged children, many of whom have substantial disabilities in both receptive and expressive language function relative to expectations of a middle-class Anglo-American culture.

The purpose of a study by Bradley (1967) was to investigate selected characteristics, academic performance, personal problems, and successes of Negro undergraduates in seven formerly all-white Tennessee colleges and universities. In addition to course grades, personal and social data were collected on 583 students over a two-year period by means of interviews and a student questionnaire. One result is pertinent for reporting here. The multiple regression equation for best predictions of grade point average (GPA) includes these variables *in this order:* (1) high school GPA, (2) a confidence in ability factor, (3) the American College Testing Program (ACTP) social studies score, and (4) a morale factor. The multiple R predicting college grades was .61, with a standard error of estimate of .55 (one half the difference between two letter grades, as C and B). Interestingly, Bradley found that no ACT score other than that for social studies added any significant increase. In Bradley's words: "The ACT scores in English and math cannot be used as a basis for predicting the academic success of the Negro

*The authors write: "Time limits of convergent tests favor the time-conscious middle-class culture."

students in the same way that they are used to predict college success for privileged white students."

Boney (1966) studied 104 Negro boys and 118 Negro girls in Grade 12 in a Port Arthur, Texas, high school. The *School and College Ability Test* (SCAT) had been given in Grade 8. Three subtests from the *Differential Aptitude Tests* (DAT) were administered at the end of Grade 12, concurrent with the computation of the grade point average (GPA). A multiple correlation of .80 for boys and .82 for girls resulted when the predictors of junior high school grade point average, the *Sequential Tests of Educational Progress* (STEP) in Language and Social Studies, the *California Test of Mental Maturity*. and the three DAT subtests were combined. Because 97 percent of the parents were unskilled laborers, there was little discrimination in socioeconomic status (SES) and SES did not become part of the regression equation. Boney concluded that "Negro students are as predictable as other groups" and that "prediction could be made in junior high school."

Two recent studies of the predictive validity of college admissions tests with Negro candidates appear to bear out the research findings of the College Entrance Examination Board presented earlier in this section.

Wilson (1969) reported a study undertaken by the College Research Center in order to facilitate the efforts of a group of eight highly selective liberal arts colleges for women to evaluate the progress of black students enrolled at the time and to develop rationales for extending educational opportunity to members of disadvantaged minority groups. The study focused on (a) selected characteristics of black women who entered member colleges of the College Research Center in 1965, 1966, and 1967, and (b) the correlational validity of standard admissions criteria for predicting college grades.

Black students entering CRC-colleges during the study, themselves a select group, differed from their classmates in a variety of educationally relevant ways—in socioeconomic background, career orientations, perceived purposes of college, educational plans, and attitudes, and in level of performance on standard admissions variables (measures of academic aptitude, SAT Verbal and Mathematical), scores on College Board Achievement Tests, and in secondary school standing. The findings of the study suggest that, despite such differences, forecasts of freshman-year academic performance are likely to be at least as accurate for black students as for their white classmates. There is, moreover, some evidence that predictions made on the basis of standard formulas may tend to overestimate the first-year performance of black students in the several colleges studied.

"It is commonly assumed that scholastic aptitude tests are biased against culturally different or disadvantaged students . . . but it is important to know whether they have useful validities for predicting relative criteria for such students." So wrote Munday (1965), who studied the predictive value of the American College Testing Program (ACTP) for 1,658 students in five 4-year Negro colleges in four different southern states. Munday employed five separate criteria (college English, mathematics, science, social studies, and overall averages). He found that the multiple R's derived from optimally weighting four high school grades in each category was lower than the multiple R's derived from the optimal weighting of the four ACTP tests. The latter R's gave predictions of college grades that were as good for the Negro colleges as for all colleges using the ACT service.

Munday described his findings as being consistent with those from other studies, that is, that grades for socially disadvantaged students are generally as predictable as grades for other students using standardized measures of academic ability. In Munday's words: "If such tests are culture-bound, as seems likely, this feature does not appear to detract from their usefulness as predictors of academic success."

MEXICAN-AMERICAN STUDIES

In one of a series of studies investigating the possible bias of testing Spanish-speaking children in English, Davis and Personke (1968) gathered evidence concerning the effects of administering the *Metropolitan Readiness Tests* (MRT) in English and Spanish to 88 Spanish-speaking children in their first school year in a South Texas city. Fifty-three of the children were enrolled in pre-first grade sections, or "readiness classes" designed for children deficient in the English language; 35 of the children were in regular first-grade sections. Early in the school year, the Spanish version of the MRT, with published test directions in English translated into South Texas colloquial Spanish, was administered to all of the children by the same individual, and the English version, according to school practices, by the classroom teachers. Contrasts of mean differences on subtest and total scores on the two modes of test administration yielded mostly nonsignificant differences. The children performed at a significantly higher level on the subtests on Word Meaning when the test was administered in Spanish; on the subtests on Alphabet and Numbers, however, significant differences favored the administration of the test in English. The findings did not show that administration of the MRT in English rather than Spanish resulted in any inadequate assessment of and substantial testing bias against Spanish-speaking children.

As a second phase of this study, Personke and Davis (1969) administered the *Metropolitan Achievement Tests* (MAT) in May to the first graders who had participated in the earlier testing with the MRT. The total score on the English administration of the MRT was a significantly better predictor of performance on the Word Knowledge subtest of the MAT than was the total score on the Spanish administration. For the other two subscores on the MAT, Word Discrimination and Reading, the English administration of the MAT yielded higher, but not significantly different, coefficients of correlation than the Spanish administration did. Of 12 comparisons made between the subtests of the MRT (English and Spanish versions) and the three scores on the MAT, six differences were statistically significant, and these differences divided themselves equally between the English and Spanish administrations. The administration of the MRT in English rather than in the children's native Spanish apparently did not result in test bias for these children.

While the results of this research are interesting and impressive, one wonders how any other outcomes could have been anticipated. If children are being taught to read English, then their readiness to learn should be best assessed in terms of their ability to cope with the English language; and the greater that ability, the greater the amount of progress in reading achievement to be expected.

Karabinus and Hurt (1969) described the results of the revised *Van Alstyne Picture Vocabulary Test* given to 535 six-year-old Mexican-American children attending poverty-qualifying schools in Tucson, Arizona. Spearman-Brown, Kuder-Richardson, and test-retest reliability coefficients for the scores of the Mexican-American children ranged from .76 (Kuder-Richardson) to .87 (test-retest), as compared with .71 (Spearman-Brown) for the general norming population. Concurrent validity coefficients with the *Stanford-Binet Intelligence Scale,* the *Wechsler Intelligence Scale for Children,* and the *Metropolitan Readiness Tests* were above .60. While the Van Alstyne test was judged to be both reliable and valid for the measurement of mental ability of these Mexican-American children, the mean mental age for the Tucson group was so much lower than that of six-year-old children in the population used for norming (33.4 as opposed to 44 to 47 months) that a normalized frequency distribution of raw scores showing corresponding percentile ranks was developed for use with the Mexican-American children rather than the percentile ranks for IQ scores provided in the manual. It was suggested that the special norms might be useful when measuring other culturally disadvantaged children.

Morper (1967) studied the relationship between certain predictive variables and achievement measures for Spanish-American and Anglo ninth graders in Oklahoma. To 50 children of each ethnic group he administered the *Wechsler Intelligence Scale for Children* (WISC), the *Lorge-Thorndike Intelligence Tests,* and the *School and College Ability Test* (SCAT) as predictive measures. Achievement measures included teacher marks in English, mathematics, and science and the *Metropolitan Achievement Tests.* For the Spanish-American group, neither the WISC nor the Lorge-Thorndike IQ's correlated at the 5 percent level of significance with scores on the MAT; while for the Anglo group, all three predictor variables correlated satisfactorily with the MAT scores. With teacher marks as criterion variables, the correlations for all predictive variables were significant for both ethnic groups. The greatest differences between the Spanish-American and Anglo groups were observed when reading ability and comprehension were most involved in the obtaining of a measurement, the difference being in favor of the Anglo group.

Kimball (1968) studied parent and family influences on the academic achievement of Mexican-American students. His population included 1,457 Grade 9 students from eight junior high schools, 899 Mexican-Americans and 558 Anglos. Twenty-three variables were tested for association with (1) school marks, (2) achievement test scores, and (3) general ability. Parental educational aspiration for their child was significantly related to all achievement variables and was more strongly related to achievement than were personal identity, background, family structure, social status, and ethnic status. Just below parent influence in predictive ability were percent of Anglos in the school, socioeconomic status, father's education, family intactness, family birth in Mexico, grandparents' residence, and birthplace of child. Sex, age, birth order in family, and family size were of little consequence. A comparison of Mexican-American and Anglo patterns of relationship between achievement and these independent variables was found by Kimball to indicate more overall differences than similarities.

Chandler and Plakos (1969) of the Mexican-American Education Project conducted an investigation to determine whether certain Mexican-American students belonged in Educable Mentally Retarded (EMR) classes or whether a language barrier prevented them from being assessed properly as to their native abilities to perform cognitive tasks. Their sample included 47 students of Mexican descent, with a problem in using the English language, in grades 3 through 8 in two school districts, an urban and a rural district, in different geographical areas. The Spanish version of the *Wechsler Intelligence Scale for Children*

was administered and scores interpreted in terms of norms developed in Puerto Rico. (Because this version was in Puerto-Rican Spanish, some items had to be reworded and some changes made in the key.) The IQ's so obtained were compared with previous IQ's based on a test not identified. The mean IQ gain was 12.4, with 44 of the 47 students scoring higher on the Spanish WISC. The median IQ was 83, as compared with a median IQ of 70 on the test administered earlier. Only 9 of the 47 scores were below the cutoff IQ of 75 for EMR classes when the Spanish WISC was given.

Of interest to note here is an experiment conducted by Palomares and Johnson (1966) that demonstrated the crucial role played by the psychologist in the over-representation of Mexican-American children, or, for that matter the overrepresentation of children of any minority group, in EMR classes. Palomares and Johnson each tested and interviewed approximately 35 Mexican-American children, ages 7 to 14 years, who had been recommended for EMR class placement. After testing the children with the *Wechsler Intelligence Scale for Children* (WISC), the non-Spanish-speaking psychologist, Johnson, found 24 of his 33 students, or 73 percent, eligible for EMR classes, while the Spanish-speaking psychologist, Palomares, recommended that only nine of his 35 students, or 26 percent, be placed in EMR classes. Clearly examiners, as well as tests, can differ even when the students tested are similar and the tests used, the same. There is little doubt but that a larger scale experiment would result in similar findings. Incidentally, both examiners averaged IQ estimates of 95 on the *Goodenough-Harris Draw-a-Man* and *Draw-a-Woman Test* for children on subsamples of 25 for whom the WISC total IQ's averaged 70 and 75, respectively.

Metfessel (1965) studied attitude and creativity factors related to achieving and nonachieving disadvantaged youth, largely Mexican-American. He found the *Individual Tests of Creativity* to be considerably superior in predicting the academic behavior generally, and of Mexican-Americans particularly, than traditional measures of intellect and scholastic aptitude. Correlations of the scores of these creativity tests with grade point averages were ranging from .39 to .49 at the time Metfessel reported. The *Inventory of Self Appraisal* and the *Meaning of Words Inventory,* two relatively independent tests of the achievement motive, were correlating between .36 and .44 with grade point average. Metfessel concluded that the results appeared to indicate that "the above three tests combine to produce a potent unified approach to forecast student achievements."

The eight Mexican-American studies briefly annotated here cover thinly the same general issues treated more fully for blacks and whites of low socio-economic status in the preceding sections. The added feature is the foreign language component; ghetto children suffer language handicaps, but nothing quite as "wrong" as a wholly different language base. The Palomares-Johnson difference of interpretation of essentially the same low performance on individual tests is an echo of the Kariger (1962) finding reported in the previous section that personal judgment compounds the ethnic separation produced by objective measurement.

MISUSES OF TESTS

Generally speaking, researchers are not studying or trying out and evaluating tests. They are studying other matters—problems, gains for compensatory programs, and the like. For the most part the tests are taken for granted as measuring instruments; in only a few cases are they questioned. That is undoubtedly why there are very few investigations of how well a test works—how valid it is—with specific differentiated groups. The published nationally standardized test is often accepted uncritically and/or simply used as the best available instrument for the purpose at hand.

Beyond the general acceptance of the test as "it," the search of the literature has uncovered some rather serious misuses of tests—using certain tests inappropriately, making comparisons across different tests, and reading into the test results more than the author and publisher intended. The *Peabody Picture Vocabulary Test* has been particularly misused. This easy-to-give test seems to be widely accepted as a good measure of general intelligence rather than offering an estimate (only) of verbal intelligence. It is frequently used with culturally deprived children with very limited vocabularies and the results compared with those of the norms group. Its use as a screening device is justified—nothing more.

Among other instances of misuse are these, which were written down as noted in reading the many studies abstracted for this report. The presence of a few such studies in this report is noted incidentally.

- —Assuming that a test designed for gifted children of one age is suitable, then, for use with older children with limited backgrounds. (See study by Lesser et al., pp. 71-72.)
- —More generally, assuming that a test constructed and standardized for children of a given age and/or school experience is equally valid for children of different ages and/or experience.
- —Changing some items and some credited answers, but applying the regular norms, especially with Puerto Rican and Mexican-American groups. (Noted in studies described just above).
- —Testing so early in preschool programs, in order to get a pretest base when improvement is to be mea-

sured, that test results cannot be valid. When a child has never handled pencil or crayon, never had a book or booklet and turned pages, never followed group directions, never worked steadily in a self-directed situation, then group tests like the *Metropolitan Readiness Tests* cannot be valid measures. They do not measure what the tests are designed to measure because test-taking is so new and unfamiliar. The resulting scores may be purely chance, or zero, although the children may have some degree of readiness.

—Posttests after an interval of group experience and use of crayons, and so forth, can produce a more valid result. But to measure score gains from pre- to posttesting and ascribe them to the effectiveness of the program in bringing about improvement in the traits measured is not justifiable if no training for the pretesting has been given. (Several Head Start evaluations suffer from this flaw.)

—Assuming that learning ability is measured by what has been learned, using the *Peabody Picture Vocabulary Test* or even the *Stanford-Binet Intelligence Scale* with its heavy emphasis on vocabulary, or the *Wechsler Intelligence Scale for Children,* with children with limited backgrounds. The emphasis on evaluation in these early childhood programs should be on getting children ready to be taught. The emphasis should be on current achievement, rather than on "intelligence," in assigning them to learning groups.

—Failing to separate reading and oral vocabulary in English from the appraisal of learning ability. Failure to use other than English-language tests for Mexican-American children, and then classifying low scoring pupils as mentally retarded, is a clear example. (Noted earlier).

—Doing studies with very small numbers of students. In some studies, no tests of significance have been made and, if they had been, hardly any significant (meaningful) results could have been obtained because of the tremendous differences in score that would have been required. Many findings of "no significant difference" are attributable to the small numbers of cases involved.

—Failing to follow through for two, three, four years, or more. The lack of longitudinal studies is distressing. It is little wonder that the longitudinal study of the culturally deprived in compensatory programs, being conducted under the auspices of Educational Testing Service for the U. S. Office of Education—from age three to Grade 3—has been so widely hailed. There are no others like it.

—Interpreting scores of *individuals* on short subtests when the reliability estimates, simply because of the length of the tests, make it impossible to trust the results of comparisons. Comparison of means for groups on the same data would be quite permissible because group means are often quite reliable enough for such purposes.

—Comparing reliability coefficients without reference to differences in range of scores.

—Treating different measures of learning ability as though the results on them were comparable. Often, no attention is paid to what the test is measuring, that is, to its content. Thus, the *Goodenough-Harris Draw-a-Man* and the *Peabody Picture Vocabulary Test* are often treated along with the Stanford-Binet as though they were equivalent and similar measures. Results on group pencil-and-paper tests of mental ability cannot be treated as equivalent to the results from individual testing.

—Attaching the same importance to predictive validity without intervention (in the form of compensatory training) as with it. When a minimum amount of intervention is used, predictive validity is an indicator of the usefulness of preliminary information; when substantial intervention is attempted, predictive validity is no longer subject to such simple interpretations. Successful intervention involves defeating predictions of failure.

SUMMARY AND CONCLUDING REMARKS

Just as much of the research on ability grouping has failed to produce conclusive findings regarding the advantages (and the disadvantages) of such grouping, in like manner much of the research on the testing of the culturally limited has failed to produce conclusive findings regarding either the validity of the tests for the use being made of them or the validity of the interpretations of the test results for such students.

As long ago as 1964, Fishman et al. prepared a set of "Guidelines for Testing Minority Group Children." The reader may be referred to that source for a compact summary of the major issues.

The discussion in this section has taken particular account of their first two major points regarding the importance of any differences found in reliability and predictive validity when the same instruments are used to evaluate minority and majority group children. Notice has been taken at several junctures that (1) reliability of a test is often equally great for minority as for majority groups, and (2) predictive validity is often as high for minority or mixed groups as for majority groups. In fact, instances have been reported in which predictive equations based on majority groups overpredict the subsequent academic achievement of minority students, thereby "favoring" the minority groups at choice points such as college admission or ability group assignment.

The discussion proceeds farther, however, to consideration of factors that affect both measures taken at the initial point of prediction and the later "final" point of assessing achievement. It is here that doubt and confusion remain. Equally low effort and accomplishment at both points will contribute positively to predictive validity. Does this lack of effort on tests at both points, a failure to organize oneself for the ultimate in competitive effort, constitute a fundamental defect requiring remediation? Does modern life essentially require this competitive effort? If so, can it be learned? Meanwhile, what procedures can be adopted to keep these modifiable traits from unduly influencing initial measures? Can we turn to foreign students for a cue? Must we allow practically unlimited time for initially slow-paced children so they can take their time interpreting questions, reading and "translating" multiple-choice options, carrying through problem-solving operations?

Also, can we accept as a crucial goal of modern education the separation of essential objectives basic to success in school learning and later in employment from what have been considered marks of the educated person? If so, we may be able to foster affective development of minority children and thereby indirectly their cognitive development.

REFERENCES

Abramson, Theodore. "The Influence of Examiner Race on First-Grade and Kindergarten Subjects' Peabody Picture Vocabulary Test Scores," *Journal of Educational Measurement,* 6:241-46, Winter 1969.

Alzobaie, Abdul Jabil and Newton S. Metfessel and William B. Michael. "Alternative Approaches to Assessing the Intellectual Abilities of Youth from a Culture of Poverty," *Educational and Psychological Measurement,* 28:449-55, Summer 1968.

Anastasi, Anne. "Culture-Fair Testing." *Educational Horizons,* 48:26-30, Fall 1964.

Beidler, Anne E. "The Effects of the Peabody Language Development Kits on the Intelligence, Reading, Listening, and Writing of Disadvantaged Children in the Primary Grades," Bethlehem, Pennsylvania: Lehigh University, 1968. Abstract: *Dissertation Abstracts* 29:3760-A, May 1969.

Boney, J. Don. "Predicting the Academic Achievement of Secondary School Negro Students." *Personnel and Guidance Journal.* 44:Part 2. 700-03. March 1966.

Bradley, Nolen E. "The Negro Teacher-Graduate Student: Factors Relative to Performance in Predominantly White State Colleges and Universities in Tennessee," *Journal of Negro Education,* 36:15-23, Winter 1967.

Brazziel, W. F. and Mary Terrell. "An Experiment in the Development of Readiness in a Culturally Disadvantaged Group of First Grade Children." *Journal of Negro Education,* 31:4-7, Winter 1962.

Buchanan, Richard G. "The Effect of Cultural Deprivation on Test-Taking Approach as Indicated by Response Style to Multiple-Choice Questions." New York: New York University, 1968. Abstract: *Dissertation Abstracts* 29:2994-A, March 1969.

California Elementary School Administrators Association. "A Study of Socio-Economic Status and School Achievement," 1962. (c/o California Teachers Association, 1705 Murchison Drive, Burlingame. California, 94010).

Campbell, Joel. "Testing of Culturally Deprived Groups." Unpublished Research Bulletin, RB-64-34. Princeton, New Jersey: Educational Testing Service, June 1964.

Carroll, John B. "Possible Directions in which College Board Tests of Abilities and Learning Capacities Might Be Developed." In *Report of the Commission on Tests. II. Briefs.* New York: College Entrance Examination Board, 1970. pp. 1-12.

Carver, Ronald P. "An Experiment that Failed: Designing an Aural Aptitude Test for Negroes." *College Board Review,* 70:10-14, Winter 1968-69.

Carver, Ronald P. "The Questionable Uniqueness of a Newly Developed Listening Test." *American Educational Research Journal,* 5:728-30, November 1968.

Carver, Ronald P. "Use of a Recently Developed Listening Comprehension Test to Investigate the Effect of Disadvantagement upon Verbal Proficiency." *American Educational Research Journal,* 6:263-70, March 1969.

Cattell, Raymond B. "Theory of Fluid and Crystalized Intelligence: A Critical Experiment." *Journal of Educational Psychology,* 54:1-22, February 1963.

Chandler, John T. and John Plakos. "Spanish Speaking Pupils Classified as Educable Mentally Retarded. California State Department of Education: Mexican-American Education Research Project." *Integrated Education,* 7(6):28-33, November-December 1969.

Cleary, T. Anne. "Test Bias: Prediction of Grades of Negro and White Students in Integrated Colleges." *Journal of Educational Measurement,* 5:115-24, Summer 1968.

Cleary, T. Anne and Thomas L. Hilton. "An Investigation of Item Bias." *Educational and Psychological Measurement,* 28:61-75, Spring 1968.

College Entrance Examination Board. *THE COLLEGE BOARD ADMISSIONS TESTING PROGRAM: A Technical Report on Research and Development Activities Relating to the Scholastic Aptitude Test*

and Achievement Tests. William H. Angoff, ed. New York: College Entrance Examination Board (in press).

Commission on Tests, David V. Tiedeman, chairman. *Report of the Commission on Tests: I. Righting the Balance. II. Briefs.* New York: College Entrance Examination Board, 1970.

Davis, Allison, chairman, and Kenneth Eells, Robert J. Havighurst, Virgil E. Herrick, and Ralph W. Tyler. *Intelligence and Cultural Differences,* A Study of Cultural Learning and Problem-Solving. Chicago: University of Chicago Press, 1951.

Davis, Elred Dennis. "Test Performance in Communication Skills of Pupils Attending Schools in Disadvantaged Areas of Knoxville (1965-67)." Knoxville, Tennessee: University of Tennessee, 1968. Abstract: *Dissertation Abstracts* 29:3868-A, May 1969.

Davis, O. L., Jr., and Carl R. Personke, Jr. "Effects of Administering the Metropolitan Readiness Test in English and Spanish to Spanish-speaking School Entrants." *Journal of Educational Measurement,* 5:231-34, Fall 1968.

Dowd, Gerald John. "Sex and Race Differences in the Effectiveness of Various Composite Predictors of Initial Reading Success and the Relationship of Children's Self-Perceptions to Initial Reading Success." Jamaica, New York: St. John's University, 1968. Abstract: *Dissertation Abstracts* 29:2999-A, March 1969.

Dreger, R. M. and D. S. Miller. "Comparative Psychological Studies of Negroes and Whites in the United States: 1959-1965." *Psychological Bulletin* (Monograph Supplement 70, No. 3, Part 2), 1968.

Dubin, Jerry A. and Hobart Osburn. "Speed and Practice: Effects on Negro and White Test Performance." *Journal of Applied Psychology,* 53:19-23, February 1969.

Eagle, Norman and Anna S. Harris. "Interaction of Race and Test on Reading Performance Scores." *Journal of Educational Measurement,* 6:131-35, Fall 1969.

Eells, Kenneth and Allison Davis, chairman, Robert J. Havighurst, Virgil E. Herrick, and Ralph W. Tyler. *Intelligence and Cultural Differences,* A Study of Cultural Learning and Problem-solving. Chicago: University of Chicago Press, 1951.

Feldmann, Shirley C. "Predicting Early Success." Paper read at conference of International Reading Association in Detroit, May 1965.

Fishman, Joshua, chairman, and Martin Deutsch, Leonard Kogan, Robert North, and Martin Whiteman. "Guidelines for Testing Minority Group Children." Prepared by a Work Group of the Society for the Psychological Study of Social Issues. *Journal of Social Issues,* 20:129-45, April 1964.

Flaugher, Ronald L. "Testing Practices, Minority Groups, and Higher Education: A Review and Discussion of the Research." Unpublished Research Bulletin, RB-70-41. Princeton, New Jersey: Educational Testing Service, June 1970.

Goldstein, Leo S. and Alan R. Coller, John Dill, and Howard S. Tilis. "The Effect of a Special Curriculum for Disadvantaged Children on Test-Retest Reliabilities of Three Standardized Instruments." *Journal of Educational Measurement,* 7:171-74, Fall 1970.

Green, Robert L. and Louis J. Hoffman. "A Case Study of the Effects of Educational Deprivation on Southern Rural Negro Children." *Journal of Negro Education,* 34:327-41, Summer 1965.

Harris, Albert J. and Robert J. Lovinger. "Longitudinal Measures of the Intelligence of Disadvantaged Negro Adolescents." *School Review,* 76:60-66, March 1968.

Hills, J. R. and J. C. Klock and S. Lewis. *Freshman Norms for the University System of Georgia, 1960-62.* Atlanta, Georgia: Georgia Regents of the University System, 1963.

Hughes, Robert B. and Ken Lessler. "A Comparison of WISC and Peabody Scores of Negro and White Rural School Children." *American Journal of Mental Deficiency,* 69:877-80, May 1965.

Karabinus, Robert A. and Maure Hurt, Jr. "The Van Alstyne Picture Vocabulary Test Used with Six-year-old Mexican-American Children." *Educational and Psychological Measurement,* 29:935-39, Winter 1969.

Kariger, Roger H. "The Relationship of Lane Grouping to the Socioeconomic Status of the Parents of Seventh-Grade Pupils in Three Junior High Schools." East Lansing, Michigan: Michigan State University, 1962. Abstract: *Dissertation Abstracts* 23:4586, June 1963.

Kendrick, S. A. and Charles L. Thomas. "Transition from School to College" in *Education for the Disadvantaged,* Edmund Gordon, issue ed. *Review of Educational Research,* 40:151-79, February 1970.

Kennedy, Wallace A. and Vernon Van De Riet, and James C. White, Jr. "A Normative Sample of Intelligence and Achievement of Negro Elementary School Children in the Southeastern United States." *Monograph of the Society for Research in Child Development,* 28:6, No. 90. Lafayette, Indiana: Child Development Publications for the Society for Research in Child Development, 1963.

Kimball, William L. "Parent and Family Influences on Academic Achievement Among Mexican-American Students." Los Angeles, California: University of

California, Los Angeles. Abstract: *Dissertation Abstracts* 29:1965-A, December 1968.

Kneif, L. M. and J. B. Stroud. "Intercorrelations among Various Intelligence, Achievement, and Social Class Scores." *Journal of Educational Psychology,* 50:117-20, June 1959.

Lambert, Nadine M. "The Present Status of the Culture Fair Testing Movement." *Psychology in the Schools,* 1:318-30, July 1964.

Lesser, Gerald S. and Gordon Fifer and Donald H. Clark. "Mental Abilities of Children from Different Social-Class and Cultural Groups." *Monograph of the Society for Research in Child Development,* 30:4, No. 102. Chicago: University of Chicago Press, 1965.

Lo Monaco, Leon John. "Response Levels of Disadvantaged Ninth-Grade Boys to Both Standard and Oral-Visual Administrations of Two Vocationally Relevant Instruments." New York: New York University, 1968. Abstract: *Dissertation Abstracts* 29:3004-A, March 1969.

Lorge, Irving. "Difference or Bias in Tests of Intelligence." Paper presented at 1952 Invitational Testing Conference of Educational Testing Service. In *Testing Problems in Perspective,* Anne Anastasi, ed. Washington, D. C.: American Council on Education, 1966, pp. 465-71.

Lucas, Charles M. "Survey of the Literature Relating to the Effects of Cultural Background on Aptitude Test Scores." Unpublished Research Bulletin, RB-53-13. Princeton, New Jersey: Educational Testing Service, June 30, 1953.

Machover, Solomon. "Cultural and Racial Variations in Patterns of Intellect; Performance of Negro and White Criminals on the Bellevue Adult Intelligence Scale." *Teachers College Contributions to Education.* No. 875, 1943; also *Teachers College Record,* 45:52-54, October 1943 (summary).

Metfessel, Newton S. "An Investigation of Attitudinal and Creativity Factors Related to Achieving and Nonachieving Culturally Disadvantaged Youth." Project Potential, USOE Cooperative Research Project No. 2615. Los Angeles, California: University of Southern California, 1965.

Mitchell, Blythe C. "Predictive Validity of the *Metropolitan Readiness Tests* and the *Murphy-Durrell Reading Readiness Analysis* for White and Negro Pupils." *Educational and Psychological Measurement,* 27:1047-54, Winter 1967, Part II.

Morper, Jack. "An Investigation of the Relationship of Certain Predictive Variables and Academic Achievement of Spanish-American and Anglo Pupils in Junior High School." Stillwater, Oklahoma: Oklahoma State University, 1966. Abstract: *Dissertation Abstracts* 27:4051-A, June 1967.

Munday, Leo. "Predicting College Grades in Predominantly Negro Colleges." *Journal of Educational Measurement,* 2:157-60, December 1965.

Orr, David B. and Warren R. Graham. "Development of a Listening Comprehension Test to Identify Educational Potential Among Disadvantaged High School Students." *American Educational Research Journal,* 5:167-80, March 1968.

Palomares, Uvaldo Hill and Laverne C. Johnson. "Evaluation of Mexican-American Pupils for Educable Mentally Retarded Classes." *California Education,* 3(8):27-29, April 1966.

Pelosi, John William. "A Study of the Effects of Examiner Race, Sex, and Style on Test Responses of Negro Examinees." Syracuse, New York: Syracuse University, 1968. Abstract: *Dissertation Abstracts* 29:4105-A, May 1969.

Personke, Carl R., Jr., and O. L. Davis, Jr. "Predictive Validity of English and Spanish Versions of a Readiness Test." *The Elementary School Journal* 70:79-85, November 1969.

Pettigrew, Thomas F. *A Profile of the Negro American.* Princeton, New Jersey: D. Van Nostrand Company, Inc., 1964.

Roberts, S. Oliver. *Studies in Identification of College Potential.* Nashville, Tennessee: Fisk University, Department of Psychology, 1962. Mimeographed.

Roberts, S. Oliver and E. P. Crump, Ann E. Dickerson, and C. P. Horton. "Longitudinal Performance of Negro-American Children at Five and Ten Years on the Standford Binet." Paper read at annual meeting of American Psychological Association, September 1965.

Santos, Beatriz N. "Special Achievement Testing Needs of the Educationally Disadvantaged." Iowa City, Iowa: University of Iowa, 1967. Abstract: *Dissertation Abstracts* 28:2567-A, January 1968.

Shuey, Audrey. *The Testing of Negro Intelligence,* 2nd Edition. New York: Social Science Press, 1966.

Society for the Psychological Study of Social Issues. "Guidelines for Testing Minority Group Children." Prepared by a Work Group, Joshua Fishman, chairman. *Journal of Social Issues,* 20:129-45, April 1964.

Stanley, J. C. and A. C. Porter. "Correlation of Scholastic Aptitude Test Scores with College Grades for Negroes versus Whites." *Journal of Educational Measurement,* 4:199-218, Winter 1967.

Temp, George. "Test Bias: Validity of the Scholastic Aptitude Test for Blacks and Whites in Thirteen Integrated Institutions." Princeton, New Jersey: Educational Testing Service, 1971.

Turnbull, William W. "Influence of Cultural Background on Predictive Test Scores." Paper presented at 1949 Invitational Testing Conference of Educational Testing Service. In *Testing Problems in Perspective,* Anne Anastasi, ed. Washington, D. C.: American Council on Education, 1966, pp. 458-64.

Weiner, Max and Shirley Feldmann. "Validation Studies of a Reading Prognosis Test for Children of Lower and Middle Socio-economic Status." *Educational and Psychological Measurement,* 23:807-14, Winter 1963.

Wilson, Kenneth R. "Black Students Entering CRC Colleges: Their Characteristics and Their First-Year Academic Performance." Research Memorandum 69-1. Poughkeepsie, New York: College Research Center, April 15, 1969. Mimeographed.

IV. ALTERNATIVE STRATEGIES TO ABILITY GROUPING

The research into the procedures for the use of tests in grouping students for learning has provided limited information. This research has been described in earlier sections of this report as generally inconclusive, with the learning environment uncontrolled and the affective domain de-emphasized. There is real need for a well-designed major program of longitudinal studies, including multivariate and covariate analyses, with consideration of the learning environment, in which the student's development is evaluated against criteria involving the cognitive, performance, and affective domains (Anderson, 1969). However, during the years required for such studies, certain helpful practices for the use of tests in the learning situation have been identified and can be described. The practices are concurred in by authorities from the fields of education and psychometrics.

INDIVIDUALIZED INSTRUCTION

The purpose frequently stated for grouping children in learning situations is to provide for individual differences. In this subsection, selected procedures are discussed for test utilization and the realization of individualized instruction.

Perhaps individualized instruction has as many definitions as there are "authorities" defining the term. Individualized instruction is herein thought of as a process of designing the curriculum for the individual (Goodlad, 1966; Rasmussen, 1968). In the process we would start by developing rapport with the student. As rapport is established, the teacher initiates an effort to define the student's characteristics. If not initially, as soon as feasible, tests and measures should be utilized by a competent person to assist in the definition of the student's characteristics. As the student enters school, for example, the tests might well include *individual* intelligence tests and/or reading measures.

After the teacher has established rapport with and gained a knowledge of the student, she is in a position to discuss objectives with the student. The objectives are mutually agreed upon and become those of the student. The curriculum content is selected by the teacher to support the student's objectives. The content includes relevant and realistic aspects of the cognitive, performance, and affective domains.

The student progresses at his rate in the mastery of the identified curricular content. It is emphasized that the student progresses at *his rate to mastery.* The mastery is normally determined in part, if not totally, by tests. The tests measure achievement and performance, and sample curricular content behaviors. The purpose of the testing is to establish mastery and readiness for the next curricular topic. In the event that the student has not mastered a given topic, he is not failed but continues to study the topic until mastery is obtained.

The procedures, materials, and methods used to guide the student in learning the content are individualized for the student (Glaser, 1966; Lindvall and Cox, 1969). In that the measures of cognitive processes and styles are in preliminary stages of development, they are not currently dependable for this purpose. Rather, the teacher should observe, both informally and systematically, the means whereby the student learns, and proceed to guide the student on a pragmatic basis.

* * * * *

Now that we have individualized instruction, is it possible to group students for learning? Four possible procedures are suggested. They are not exhaustive of all possible procedures. They are judged, in the light of the findings of the preceding sections, to be the most promising.

HETEROGENEOUS GROUPING

An important part of what children learn is obtained directly from other children who know things that they do not know. This may be furthered by planned heterogeneous grouping which involves the bringing together of students who deviate extensively on a given variable. For example, in an elementary school social science class a topic for discussion might be the State of California. The student's knowledge of the state is the variable. Some student might have lived or visited in the state and observed a great amount of realistic information pertaining to the state. A group is formed consisting of those knowledgeable students and those desiring to learn about the state. In this instance we have an "ad hoc" heterogeneous group. The knowledgeable members have an opportunity to gain in leadership and communication skills through instruction of the others. The others, with guidance, are motivated to learn that which their peers know.

Heterogeneous grouping of this nature is practiced in the non-graded school. Children assigned in a non-graded school vary considerably in age, experience, and knowledge. The heterogeneity is planned so that the children can learn from each other.

Heterogeneous grouping of the more common variety, putting together children in unselective fashion, may achieve the same effect if the teacher remains alert to opportunities to promote exchange of ideas, information, and skills in diverse groups. The key is to stimulate the desire to share novel information, rather than promoting headlong competition.

STRATIFIED HETEROGENEOUS GROUPING

The illustration just cited presents a clear case for the values of heterogeneous grouping. But let us consider another situation commonly faced in elementary schools in which it has been customary to teach classes of 30 children or so in self-contained classrooms where the 30 children stay with the same teacher in the same room for practically the entire day. Suppose we accept the criticism of those who argue for homogeneous ability grouping to reduce the span of achievement in each classroom, yet are even more attentive to the criticism of those who argue against homogeneous grouping of whole classrooms because of the stigma this places on those in the average and low groups while giving the high groups an unwholesome feeling of general superiority. Can these views both be accepted in a plan of organization of classrooms that has its own peculiar advantages? It has been done.

In Baltimore, a fundamental plan of organization recommended as an alternative that meets these requirements* may be called a plan of "stratified heterogeneous grouping." Under this plan, if three classes of 30 are to be made of 90 children ready to start fifth grade, the children would be ranked in order of excellence on some composite—say, a standardized test battery most recently given—and then be subdivided into nine groups of ten each. Teacher A would be given a class consisting of the highest or first ten, the fourth ten, and the seventh ten; Teacher B would have the second, fifth, and eighth tens; Teacher C would then be given the third, the sixth, and the ninth (lowest) tens, as shown below.

Teacher A	*Teacher B*	*Teacher C*
Group 1 (1-10)	Group 2 (11-20)	Group 3 (21-30)
Group 4 (31-40)	Group 5 (41-50)	Group 6 (51-60)
Group 7 (61-70)	Group 8 (71-80)	Group 9 (81-90)

Note the several merits of this scheme. First, there is no top or bottom section; the sections overlap, so invidious comparisons between groups are minimized. Second, each class has a narrower range than the full 90 have: Teacher A has the top ten, but none of the bottom 20; Teacher C has the bottom ten, but none of the top 20; Teacher B has neither the top nor the bottom ten. Third, teachers can give special attention where it is needed without feeling unable to meet the needs of the opposite extreme: Teacher A can give a little special attention to the top ten because the bottom 20 are not in the class; Teacher C can concentrate on the bottom ten, without fear of "losing" the top 20. Fourth, each class has leaders of appropriate capability to stimulate each other in a fair competitive way while giving leadership to lower groups; note particularly that in Teacher C's class, the top group is the third ten, a group that has probably always had to play second fiddle to some in the first or second ten. Finally, no teacher has to teach the bottom group of a homogeneous plan, that mixture of disruptive, leaderless children who lack motivation and capability and make teachers like homogeneous grouping, but equally dislike to teach the slow group.

Such a method of grouping is not offered as a complete answer by itself, but as a constructive step in the right direction. It is, moreover, compatible with other special teaching arrangements like team teaching, peer tutoring, and early education.

Elementary School Guide, Baltimore Public Schools, revised edition, 1967.

TEAM TEACHING WITH FLEXIBLE GROUPING

The history of heterogeneous grouping schemes is that they do not involve an additional expenditure of funds. Our third procedure is thought to involve additional funds, especially during the implementation phase. However, the additional gains in this third procedure are judged to show a favorable cost-effectiveness trade-off.

The U.S. Office of Education has sponsored a number of efforts to develop specifications for new model elementary school systems. A total of ten such models have been developed (Stauffer and Deal, 1969). Without exception, each model, with numerous variations, has embraced the concepts of individualized instruction, mastery, and differentiated staff. The differentiated staff approach specifies various personnel categories for teachers such as aides, assistants, specialists, and the like (Allen, 1967). Each category has certain functions of prime responsibility. The team teaching staff is selected from these categories of teachers so as to satisfy the requirements of a given situation.

The team would normally contain or have readily available a specialist who would perform, or guide a competent teacher in, the diagnosis of the individual student. The specialist is trained in selecting and administering tests, interpreting test results, and defining appropriate programs of instruction. After the objectives and content are defined for the student, the task of guiding the student's learning is assigned among the team members as appropriate.

In a team, normally, there is a considerable number of staff members, say six or more, and a large class, say 100 or more. Thus, it is frequently found that a number of students have a need to learn the same tasks. Groups of such students are formed and assigned to a designated teacher for the purpose of learning the specific tasks. The grouping is informal, ad hoc, and of short duration. In a situation of this nature, the

students and teachers are paired with the task to be accomplished. Grouping in this manner promotes the effective utilization of personnel and resources, and increased learning by the individual, without the identified detrimental effect of homogeneous grouping.

STUDENT TUTORING

Tutoring of children deficient in academic skills by older children has been widely adopted within compensatory education programs. Not surprisingly, those tutored show more than normal gains over a period of instruction. What is perhaps somewhat more surprising, when older children—themselves deficient in basic skills—are paid to tutor younger children who are deficient, the gains of the tutors outstrip by far the gains of the tutored!

Cloward (1967) reports a study in which children of junior high school grade status, who were two or more years retarded in reading, as measured by grade scores on standardized reading test, were paid $1.25 per hour to tutor deficient fourth-grade children of similar ethnic background (Caucasian, Puerto Rican, Negro). The program was conducted over an academic year after the tutors had undertaken a period of preparation (also on paid time) for their teaching chores. The psychodynamics of the tutor growth is worth spelling out rather fully.

First, these older students, who had experienced the constant role of failures pitied or deplored by their teachers, were now being asked, nay, even paid, to make a contribution to others. Second, in preparing for this work, they had learned the basis of the old maxim, "If you want to learn something, teach it." Third, they could see their pupils learn, as measured by daily response as well as by terminal test.

Specifically, using analyses of covariance to control for small initial differences in reading scores, Cloward found that 100 deficient readers in fourth and fifth grade who were tutored for four hours a week for 26 weeks did reliably better than 79 control children at the end of that period, reversing somewhat the normal trend toward further retardation characteristic of their peers. Tests given five months apart showed average gains of 6 months by experimentals, 3-1/2 months by controls. During the same period, 77 tutors, who averaged 0.8 grades deficient at the start, gained reliably more than their 52 controls by 1.7 grades. Bearing in mind that grade score differences at high school level are magnified by the fact that the slope of the growth curve is decreasing, the adjusted mean difference at the end is slightly more than half a standard deviation on the score scale.

A noteworthy variation on this procedure was observed at a school in another city where it was reported to be standing operating procedure. The teacher, at junior high level in a low socioeconomic area, had a class consisting in equal parts of delinquent and mentally retarded white boys. She paired off each delinquent with a mentally retarded boy of the same age and taught the delinquent to get a new satisfaction from his ability to help and teach a mentally retarded boy. It was heartening to watch pairs of boys come forward to show what the slow learner of each pair had accomplished.

EARLY CHILDHOOD EDUCATION

At least since the 1930's, when the studies emanating from the Iowa Child Welfare Research Station (Stoddard, 1943) challenged the then accepted concept of the constancy of the IQ (Hunt, 1961) with evidence that substantial gains or losses in intellectual competence could be generated by the nature of early environmental stimulation of children, many parents from the upper socioeconomic classes have been sending their children to nursery schools. Beginning sometimes as early as age 2, these children have enjoyed intellectual stimulation in a supportive emotional climate and have emerged readier to participate in conventional schooling at age 5 or 6. In many such schools, priority has been given to affective development over intellectual stimulation. In others, however, intellectual stimulation has been an integral feature of this early education.

Currently, the debate rages about whether this early intellectual stimulation may be cast in a form that is best called early schooling, the earlier presentation of instructional stimulation ordinarily offered all comers at an approximately uniform starting point of age 6 in Grade 1. What is best done at earlier ages is still moot, but experiments with children beginning at age 5 in Kindergarten (McKee and Brzeinski, 1966; Brzeinski et al., 1967; Fortson, 1969) show conclusively effective gains from planned early schooling in Kindergarten. The Denver data reported by Brzeinski show that reliable gains from such early instruction in reading persist at least through Grade 5, with some spread to related curriculum areas. An important condition is that gains achieved in Kindergarten shall be consciously built upon in successive grades rather than being left to conventional programs for incidental forwarding; indeed, children placed in conventional classes with children beginning the learning of reading at age 6 in Grade 1 soon slip from being recognized by their teachers as advanced at that point to becoming ones less challenged by the teaching of already learned skills and eventually being not at all advanced over their peers.

Implications of these and other findings for the enhancement of learning by disadvantaged groups would appear to be that the practice of beginning

formal instruction at age 5 (with some imaginative adaptations) might well follow the established practice of the British Infant School of beginning instruction for all children at this level.

SUMMARY AND CONCLUDING REMARKS

This section concludes our report with a series of brief accounts of alternative strategies to ability grouping. These illustrations by no means exhaust the possibilities, but they constitute a set of mutually compatible strategies each of which has separate merit. Heterogeneous grouping promotes communication and peer teaching. Stratified heterogeneous grouping furthers these same goals while reducing the extreme variations in a class that complicate group instruction. Team teaching permits flexible grouping to achieve individual learning objectives. Student tutoring promotes learning by the tutors as well as by the tutored, a circumstance also furthered by stratified grouping. Early childhood education, at least from Kindergarten at age 5, can undergird a persistent gain in mastery of fundamentals. Taken together, these alternative strategies constitute a constructive challenge to the unrealized advantages and actual deleterious effects of ability grouping in the areas of scholastic achievement, affective development, and the ethnic and socioeconomic separation (isolation, deprivation) of children.

REFERENCES

Allen, D. W. "Differential Teaching Staff." Interim seminar paper presented at Stanford University, Palo Alto, California, March 1967.

Anderson, Scarvia B. "The ETS-OEO Longitudinal Study of Disadvantaged Children." In *Untangling the Tangled Web of Education,* a special symposium sponsored by the National Council on Measurement in Education. Princeton, New Jersey: Educational Testing Service, 1969. pp. 27-38.

Baltimore Public Schools. *Elementary School Guide,* revised edition. Baltimore, Maryland, 1967.

Cloward, Robert D. "Studies in Tutoring." *Journal of Experimental Education,* 36:14-25, Fall 1967.

Fortson, Laura R. "A Creative Aesthetic Approach to Readiness and Beginning Reading and Mathematics in Kindergarten." Athens, Georgia: University of Georgia, 1969. Abstract: *Dissertation Abstracts* 30:5346-A, June 1970.

Glaser, Robert. "The Design of Instruction." In *The Changing American School,* John I. Goodlad, ed. Sixty-fifth Yearbook of the National Society for the Study of Education, Part II, pp. 215-42. Chicago: University of Chicago Press, 1966.

Goodlad, John I. "Diagnosis and Prescription in Educational Practice." In *New Approaches to Individualizing Instruction,* a report of a conference in May, 1965, to mark the dedication of Ben D. Wood Hall. Princeton, New Jersey: Educational Testing Service, 1966, pp. 27-37.

Hunt, J. McV. *Intelligence and Experience.* New York: Ronald Press, 1961.

Lindvall, C. M. and R. C. Cox. "The Role of Evaluation in Programs for Individualized Instruction." In *Educational Evaluation: New Roles, New Means,* Ralph W. Tyler, ed. Sixty-eighth Yearbook of the National Society for the Study of Education, Part II, Chicago: University of Chicago Press, 1969, pp. 156-88.

McKee, Paul R. and J. E. Brzeinski. *The Effectiveness of Teaching Reading in Kindergarten.* Cooperative Research Project No. 5-0371. Denver, Colorado: Denver Public Schools, 1966.

Rasmussen, L. V. *Meeting the Critics' Demands for Quality Education—through Individualized Instruction.* A Special Report. Washington, D. C.: The National Laboratory for the Advancement of Education, 1968, pp. 12-15.

Stauffer, A. J. and T. N. Deal, eds. "Teacher Evaluation Models." *Journal of Research and Development in Education,* 2(3):3-135, Spring 1969.

Stoddard, George D. *The Meaning of Intelligence.* New York: Macmillan Co., 1943.

APPENDIX A

A NOTE ON JENSEN AND OTHER NEW DEVELOPMENTS

Because of the widespread publicity achieved by the debate over an article entitled "How Much Can We Boost the IQ and Scholastic Achievement?" by Arthur R. Jensen in the Winter, 1969, issue of the *Harvard Educational Review,* some readers may wonder at its relevance to the issue of ability grouping. Jensen suggests that some children learn better by association (rote memory) and others by fitting new learning into a conceptual framework by higher mental processes, and that the whole matter of efficient learning styles is related to genetically determined "intelligence" in which certain ethnic groups are on the average considerably better endowed than others.

The readers is referred to the considerable bibliography of critical replies in subsequent issues of the *Harvard Educational Review* and elsewhere, listed at the end of this appendix. Suffice it here to quote from Cronbach's response and add our abbreviated critique.

Cronbach (1969) says in part:

> Professor Jensen is among the most capable of today's educational psychologists. His research is energetic and imaginative. In the present paper, an impressive example of his thoroughness, I am sure every reader has had my experience of encountering valuable information in areas where he thought himself *au courant.* Unfortunately, Dr. Jensen has girded himself for a holy war against "environmentalists" and his zeal leads him into over-statements and misstatements.

Despite the merits of Jensen's research remarked by Cronbach, and admitting the dubious propriety of some of the criticism addressed to Jensen for publishing data and argument that may be used for partisan ends, his presentation suffers from faults in at least five major respects:

1. Jensen starts in journalistic style to proclaim a finding, rather than in professional style to build a convincing case.
2. Current brief and fragmented efforts at compensatory education show little effect, but it is too much to say compensatory education has failed. Efforts expended on short-term early education have produced modest gains in some instances; other experiments here and in other countries have succeeded (Brzeinski, 1967; Bloom, 1969). One might fairly add that no major effort comparable to the systematic discrimination of over three centuries against American blacks has even been attempted.
3. Traits with high heritability are often modifiable (Goldstein 1969).
4. Education's business is with a substantial modifiability. Even a correlation of .87 between monozygotic twins leaves 25 percent of the variance unaccounted for (Bloom 1969).
5. Jensen closes on a note that suggests the likelihood of his model of distinctive learning styles for variously different children without clear evidence of the likely effectiveness of different teaching styles for classroom groups. Since disadvantagedness to Jensen is an individual characteristic compounded of individual and group hereditary and environmental factors and their interactions, this can only imply responsiveness of teachers to all children with a variety of teaching styles rather than heavy dependence on one teaching style for children of each of the different learning styles. His discussion, moreover, leaves entirely out of consideration the teaching and learning that go on between children.

Other new proposals, like performance contracts and vouchering of funds to parents to let them "buy" their children's education from the best sources, are merely noted here. They are procedural rather than instructional variations. If used, it would remain for instruction to be designed as suggested here, or by more ingenious instructional plans; performance contracts and vouchering merely establish different contractual arrangements for authorizing instructional activity.

REFERENCES

Anastasiow, N. J. "Educational Relevance and Jensen's Conclusions." *Phi Delta Kappan,* 51:32-5, Summer 1969.

Bereiter, Carl. "The Future of Individual Differences." *Harvard Educational Review,* 39:310-18, Spring 1969.

Bloom, Benjamin S. Letter to the Editor. *Harvard Educational Review,* 39:419-21, Spring 1969.

Brazziel, William F. "A Letter from the South." *Harvard Educational Review,* 39:348-56, Spring 1969.

Brazziel, William F. "Perspective on the Jensen Affair." *Childhood Education,* 46:371-2, April 1970.

Brazziel, William F. "Symposium: The Jensen Controversy. I. Beyond the Sound and Fury." *Measurement and Evaluation in Guidance* 3:7-9, Spring 1970.

Brazziel, William F. and Frederick G. Brown and Howard Cameron. "Symposium: The Jensen Controversy." *Measurement and Evaluation in Guidance,* 3:7-24, Spring 1970.

Brzeinski, Joseph E. and M. Lucile Harrison and Paul McKee. "Should Johnny Read in Kindergarten?" *NEA Journal,* 56 (3):23-25, March 1967.

Brown, Frederick G. "Symposium: The Jensen Controversy. III. Review of the Past, Focus for the Future." *Measurement and Evaluation in Guidance,* 3:18-24, Spring 1970.

Cameron, Howard. "Symposium: The Jensen Controversy. II. Cultural Myopia." *Measurement and Evaluation in Guidance,* 3:10-17, Spring 1970.

Comer, J. P. "Research and the Black Backlash." *American Journal of Orthopsychiatry,* 40:8-11, January 1970.

Cronbach, Lee J. "Heredity, Environment and Educational Policy." *Harvard Educational Review,* 39:338-47, Spring 1969.

Crow, James F. "Genetic Theories and Influences: Comments on the Value of Diversity," *Harvard Educational Review,* 39:301-09, Spring 1969.

Deutsch, Martin. "Happenings on the Way Back to the Forum." *Harvard Educational Review,* 39:523-57, Summer 1969.

Elkind, David. "Piagetian and Psychometric Conceptions of Intelligence." *Harvard Educational Review,* 39:319-37, Spring 1969.

Fehr, F. S. "Critique of Hereditarian Accounts of Intelligence and Contrary Findings." *Harvard Educational Review,* 39:571-80, Summer 1969.

Goldstein, Allen C. "A Flaw in Jensen's Use of Heritability Data." *IRCD Builetin,* 5:4:7-9, Fall 1969.

Gordon, Edmund W. "Education, Ethnicity, Genetics, and Intelligence." *IRCD Bulletin,* 5:4:1, 2, 13, Fall 1969.

Gruber, Howard E. "Psychologists Comment on Current IQ Controversy: Heredity versus Environment." *IRCD Bulletin,* 5:4:6, Fall 1969.

Hirsch, Jerry. "Behavior-Genetic Analysis and Its Bisocial Consequence." *IRCD Bulletin,* 5:4:3-4, Fall 1969.

Hudson, L. "Nature, Nurture: Racialist Comeback?" *Times Educational Supplement,* 2824:33, July 4, 1969.

Humphreys, L. G. and H. P. Dachler. "Jensen's Theory of Intelligence." With Reply by Jensen and rejoinder by authors. *Journal of Educational Psychology,* 60:419-33, December 1969.

Hunt, J. McV. "Has Compensatory Education Failed: Has It Been Attempted?" *Harvard Educational Review,* 39:278-300, Spring 1969.

Jensen, Arthur R. "How Much Can We Boost the IQ and Scholastic Achievement?" *Harvard Educational Review,* 39:1-123, Winter 1969.

Jensen, Arthur R. "Jensen's Theory of Intelligence." *Journal of Educational Pyschology,* 60:427-31, December 1969.

Jensen, Arthur R. "Reducing the Heredity—Environmental Uncertainty: A Reply." *Harvard Educational Review,* 39:449-83, Summer 1969.

Kagan, Jerome S. "Inadequate Evidence and Illogical Conclusions." *Harvard Educational Review,* 39:274-77, Spring 1969.

Light, R. J. and P. V. Smith. "Social Allocation Models of Intelligence." *Harvard Educational Review,* 39:484-510, Summer 1969.

Scriven, Michael. "The Values of the Academy (Moral Issues for American Education and Educational Research Arising from the Jensen Case)." *Review of Educational Research,* 40:541-49, October 1970.

Stinchcombe, A. L. "Environment: the Cumulation of Effects." *Harvard Educational Review,* 39:511-22, Summer 1969.

Voyat, Gilbert. "IQ: God-given or Man-made?" *Education Digest,* 35:1-4, October 1969.

Zach, L. "I.Q. Test: Does It Make Black Children Unequal?" *School Review,* 78:249-58, February 1970.

APPENDIX B

Title	Publisher	Page(s)
ABILITY SELF-CONCEPT RATING (See also Drews' ABILITY SELF-CONCEPT RATING.)	Unpublished	33
AMERICAN COLLEGE TESTS (ACT)	American College Testing Program	67, 75, 76
Bills' INDEX OF ADJUSTMENT AND VALUES (See also INDEX OF ADJUSTMENT AND VALUES.)	University of Alabama	33, 34, 35
CALIFORNIA ACHIEVEMENT TESTS	California Test Bureau	11, 26, 28, 44 52, 73, 74
CALIFORNIA TEST OF MENTAL MATURITY	California Test Bureau	28, 33, 44, 76
CALIFORNIA TEST OF PERSONALITY	California Test Bureau	33, 36
CATTELL CULTURE FAIR INTELLIGENCE TEST (See also IPAT CULTURE FAIR INTELLIGENCE TEST.)	Bobbs-Merrill Co., Inc.	72, 75
Clark and Ozehosky U-SCALE (See also U-SCALE)	St. John's University Press	72
CLYMER-BARRETT PREREADING BATTERY	Personnel Press, Inc.	62
COGNITIVE HOME ENVIRONMENT SCALE (See also Ypsilanti COGNITIVE HOME ENVIRONMENT SCALE. In Radin, Norma and Hanne Sonquist *The Gale Preschool Program.* Final Report.)	Ypsilanti Public Schools, Ypsilanti, Michigan	61
COLLEGE BOARD ACHIEVEMENT TESTS	College Entrance Examination Board	67, 76
COLLEGE BOARD ADMISSIONS TESTS	College Entrance Examination Board	67
COLUMBIA CLASSROOM SOCIAL DISTANCE SCALE (In Cunningham, Ruth et al., *Understanding the Behavior of Boys and Girls.)*	Bureau of Publications, Teachers College, Columbia University	32, 34
COLUMBIA MENTAL MATURITY SCALE	Harcourt Brace Jovanovich, Inc.	71
CONCEPT OF SELF-AS-A-LEARNER SCALE (See also Drews' CONCEPT OF SELF-AS-A-LEARNER SCALE.)	Unpublished	33, 35
COOPERATIVE PRIMARY TESTS	Educational Testing Service	72
DAVIDSON-LANG CHECK LIST OF 35 TRAIT NAMES	Unpublished	34
DAVIS-EELLS GAMES	Harcourt Brace Jovanovich, Inc.	74
DIFFERENTIAL APTITUDE TESTS	Psychological Corporation	11, 76
Drews' ABILITY SELF-CONCEPT RATING (See also ABILITY SELF-CONCEPT RATING.)	Unpublished	33
Drews' CONCEPT OF SELF-AS-A-LEARNER SCALE (See also CONCEPT OF SELF-AS-A-LEARNER SCALE.)	Unpublished	33, 35

Title	Publisher	Page(s)
EMPLOYEE APTITUDE SURVEY	Psychological Services, Inc.	70
ENVIRONMENTAL PROCESS SCALE (See also Wolf's ENVIRONMENTAL PROCESS SCALE. In the author's unpublished dissertation "The Identification and Measurement of Environmental Variables Related to Intelligence," University of Chicago, 1964.)	Unpublished	61
GATES PRIMARY READING TESTS: WORD RECOGNITION, SENTENCE READING, PARAGRAPH READING	Teachers College Press Columbia University	67, 68, 72
GENERAL ANXIETY SCALE FOR CHILDREN	Yale University Press	33, 37, 38
GOODENOUGH-HARRIS DRAW-A-MAN (DRAW-A-WOMAN) TEST	Harcourt Brace Jovanovich, Inc.	78, 79
Haller OCCUPATIONAL ASPIRATION SCALE (See also OCCUPATIONAL ASPIRATION SCALE.)	Agricultural Experiment Station, Michigan State University	36
HOLLINGSHEAD AND REDLICH INDEX	Yale University Press	71
HOLLINGSHEAD TWO-FACTOR INDEX	Yale University Press	72
HUNTER APTITUDE SCALES (Adapted from the *Hunter Elementary School Admissions Test*)	Unpublished	71
I GUESS MY SCORE (In Goldberg, Miriam L. et al., *The Effects of Ability Grouping*)	Teachers College Press Columbia University	35
INDEX OF ADJUSTMENT AND VALUES (See also Bills' INDEX OF ADJUSTMENT AND VALUES.)	University of Alabama	33, 34, 35
INDEX OF SOCIAL CLASS (See also Warner's INDEX OF SOCIAL CLASS.)	Harper and Row Publishers, Inc.	45, 75
INDEX OF STATUS CHARACTERISTICS (See also Warner's INDEX OF STATUS CHARACTERISTICS.)	Harper and Row Publishers, Inc.	74
INDIVIDUAL TESTS OF CREATIVITY (See also Metfessel and Risser's INDIVIDUAL TESTS OF CREATIVITY.)	Project Potential, University of Southern California	78
INVENTORY OF ATTITUDES TOWARD JUNIOR HIGH SCHOOL	Unpublished	33
INVENTORY OF SELF-APPRAISAL (See also Metfessel's INVENTORY OF SELF-APPRAISAL.)	Project Potential, University of Southern California	78
IOWA TESTS OF BASIC SKILLS	Houghton Mifflin Co.	11, 28, 29, 44 74, 75
IOWA TESTS OF EDUCATIONAL DEVELOPMENT	Science Research Associates, Inc.	11, 28

Title	Publisher	Page(s)
IPAT CULTURE FAIR INTELLIGENCE TEST (See also CATTELL CULTURE FAIR INTELLIGENCE TEST.)	Institute for Personality and Ability Testing	69
Jordan and Thurstone's SOCIAL TRAITS RATING SCALE (See also SOCIAL TRAITS RATING SCALE.)	Unpublished	32
KUDER PREFERENCE RECORD-VOCATIONAL	Science Research Associates, Inc.	70
KUHLMANN-ANDERSON INTELLIGENCE TESTS	Personnel Press, Inc.	43
LEE-CLARK READING READINESS TEST	California Test Bureau	72
LIFE-PLANNING QUESTIONNAIRE-MODIFIED (Adapted from an instrument used in the unpublished dissertation of Hamburger, Martin, "The Relationship of Mental Ability to Occupational Choice of Adults," Teachers College, Columbia University, 1968).	Unpublished	70
LORGE-THORNDIKE INTELLIGENCE TESTS	Houghton Mifflin Co.	11, 44, 67, 74 75, 77
Maslow SECURITY-INSECURITY INVENTORY (See also SECURITY-INSECURITY INVENTORY.)	Consulting Psychologists Press, Inc.	33
MEANING OF WORDS INVENTORY (See also Metfessel's MEANING OF WORDS INVENTORY.)	Project Potential, University of Southern California	78
MENTAL HEALTH ANALYSIS SCALE	California Test Bureau	32
Metfessel and Risser's INDIVIDUAL TESTS OF CREATIVITY (See also INDIVIDUAL TESTS OF CREATIVITY.)	Project Potential, University of Southern California	78
Metfessel's INVENTORY OF SELF-APPRAISAL (See also INVENTORY OF SELF-APPRAISAL.)	Project Potential, University of Southern California	78
Metfessel's MEANING OF WORDS INVENTORY (See also MEANING OF WORDS INVENTORY.)	Project Potential, University of Southern California	78
METROPOLITAN ACHIEVEMENT TESTS	Harcourt Brace Jovanovich, Inc.	11, 68, 70, 74 75, 77
METROPOLITAN READINESS TESTS	Harcourt Brace Jovanovich, Inc.	11, 62, 63, 64 65, 72, 76, 77 79
METROPOLITAN READING TEST (See also METROPOLITAN ACHIEVEMENT TESTS.)	Harcourt Brace Jovanovich, Inc.	68, 70
MOONEY PROBLEM CHECK LIST	Psychological Corporation	33
MURPHY-DURRELL READING READINESS ANALYSIS	Harcourt Brace Jovanovich, Inc.	63, 64
N. F. E. R. PRIMARY VERBAL TEST I	National Foundation for Research in England and Wales	37

Title	Publisher	Page(s)
STEP LISTENING (See also SEQUENTIAL TESTS OF EDUCATIONAL PROGRESS.)	Educational Testing Service	68
STEP READING (See also SEQUENTIAL TESTS OF EDUCATIONAL PROGRESS.)	Educational Testing Service	68
Stott's SIX ADVERSE ADJUSTMENT POINTERS (See also SIX ADVERSE ADJUSTMENT POINTERS.)	(British Publisher Not Identified)	38
TEST ANXIETY SCALE FOR CHILDREN	Yale University Press	33, 37, 38
TEST OF ACADEMIC PERFORMANCE-READING	Houghton Mifflin Co.	75
THINKING ABOUT YOURSELF (From *Process for In-School Screening of Children with Emotional Handicaps*)	Educational Testing Service	33
U-SCALE (See also Clark and Ozehosky U-SCALE.)	St. John's University Press	72
UNIT SCALES OF ATTAINMENT	Educational Test Bureau	43
VAN ALSTYNE PICTURE VOCABULARY TEST	Harcourt Brace Jovanovich, Inc.	72, 77
Warner's INDEX OF SOCIAL CLASS (See also INDEX OF SOCIAL CLASS. In Warner, William L. and Marchia Meeker and Kenneth Eells, *Social Class in America,* a manual)	Harper and Row Publishers, Inc.	45, 75
Warner's INDEX OF STATUS CHARACTERISTICS (See also INDEX OF STATUS CHARACTERISTICS. In Warner, William L. and Marchia Meeker and Kenneth Eells, *Social Class in America,* a manual)	Harper and Row Publishers, Inc.	74
WECHSLER ADULT INTELLIGENCE SCALE (WAIS)	Psychological Corporation	69, 70
WECHSLER INTELLIGENCE SCALE FOR CHILDREN (WISC)	Psychological Corporation	72, 73, 74, 77 78, 79
Wolfe's ENVIRONMENTAL PROCESS SCALE (See also ENVIRONMENTAL PROCESS SCALE.)	Unpublished	61
~~WORD~~ RATING LIST	Unpublished	34
Ypsilanti COGNITIVE HOME ENVIRONMENT SCALE (See also COGNITIVE HOME ENVIRONMENT SCALE.)	Ypsilanti Public Schools, Ypsilanti, Michigan	61