72-3437

SMITH, Kenneth Robert, 1942-
TEACHER PREFERENCE COMPARED WITH TEACHER EFFECTIVENESS EVIDENCED BY PRODUCT MEASUREMENT: AN INDICATOR FOR TEACHER PLACEMENT IN GROUPED CLASSES.

The University of Oklahoma, Ed.D., 1971
Education, theory and practice

University Microfilms, A XEROX Company, Ann Arbor, Michigan

© COPYRIGHTED BY

KENNETH ROBERT SMITH

1971

THIS DISSERTATION HAS BEEN MICROFLIMED EXACTLY AS RECEIVED

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/3149

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

TEACHER PREFERENCE COMPARED WITH TEACHER EFFECTIVENESS EVIDENCED BY PRODUCT MEASUREMENT: AN INDICATOR FOR TEACHER PLACEMENT IN GROUPED CLASSES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF EDUCATION

BY

KENNETH ROBERT SMITH

Norman, Oklahoma

1971

TEACHER PREFERENCE COMPARED WITH TEACHER EFFECTIVENESS EVIDENCED BY PRODUCT MEASUREMENT: AN INDICATOR FOR TEACHER PLACEMENT IN GROUPED CLASSES

APPROVED BY

DISSERTATION COMMITTEE

PLEASE NOTE:

Some Pages have indistinct print. Filmed as received.

UNIVERSITY MICROFILMS

ACKNOWLEDGEMENTS

Appreciation is expressed to the four members of this dissertation committee for their understanding patience and help. It would have been difficult to find a better committee than Dr. Gerald Kidd, chairman, Dr. Jack Parker, Dr. Robert Bibens, and Dr. Lloyd Williams.

Expressions of gratitude are extended to many persons who have participated in the editing as well as consulting about the procedural design of this work. At the risk of omitting names, the following have been of such help: Dr. George T. Guess, Miss Lucile Searcy, Miss Jean Sugden, Mr. Sam Northington, Dr. George Letchworth, Dr. Donald Reynolds, and Dr. Bill Coffia.

Gratitude is expressed to the administration, faculty, and students of the Duncan, Oklahoma, Public School System without whose help the project would have been impossible. Especially to the Superintendent of Schools, Mr. Bob Brandeberry, and the Director of Elementary Education, Mr. Donald Hodge, a special note of appreciation is afforded for their efforts in making this project a reality.

Finally, to a fine wife, who has acted in all capacities from "idea generator" to secretary, a special recognition is in order. Without her backing, none of the work on this degree would have been completed.

TABLE OF CONTENTS

Chapter

I. INTRODUCTION 1

Problem
Definitions
Hypotheses
Limitations

II. REVIEW OF THE LITERATURE 11

Intelligence Quotient or Mental Age as an Indicator of Capacity to Learn
Consistency of the Intelligency Quotient
Relevant Assumptions of Past Studies
Individual Differences in Teachers
Teacher Prediction of His Effective Area

III. PROCEDURE 22

Population Description
Outline of the Procedure
Design
Method of Collecting Data
Test Instruments
School Norm
Teacher's Norm
Norm Comparison Procedure
Interview and Questionnaire
Statistical Comparisons of Indicators of Teacher Effectiveness
Summary

IV. FINDINGS 41

Development of the Means
Comparison of the Means
Determination of Effectiveness in Each Third

Results of the Questionnaire
Comparison of Indicators of Effectiveness
Summary of the Findings

V. SUMMARY, CONCLUSIONS, AND IMPLICATIONS 56

Summary of Procedures and Findings
Conclusions
Implications
Need for Further Research

BIBLIOGRAPHY 64

APPENDICES 70

LIST OF TABLES

Table		
1.	School Norms ±1 Standard Deviation	44
2.	Teacher Effectiveness, in Thirds, from Mean Tables	46
3.	Teachers' Choices on the Questionnaire . . .	48
4.	Computation of the Chi Square	52-53
5a-b.	School Norms	73-74
6a-k.	Teachers' Means	75-85

LIST OF ILLUSTRATIONS

Figure
1. Matrix for Comparison of Indicators 36
2. "Observed" Matrix 49
3. "Expected" Matrix 51

TEACHER PREFERENCE COMPARED WITH TEACHER EFFECTIVENESS EVIDENCED BY PRODUCT MEASUREMENT: AN INDICATOR FOR TEACHER PLACEMENT IN GROUPED CLASSES

CHAPTER I

INTRODUCTION

To improve class organization based on chronological age or grade level, many schools currently group children homogeneously on such bases as mental age, intelligence quotient, or achievement in academic subjects. A 1954 survey by the U. S. Office of Education revealed that approximately 50 per cent of the schools in America used some form of homogeneous grouping.[1] Because of recent extensive use of curricular innovations involving individual differences, some educators feel that a far greater percentage of schools now than in the past use some form of grouping.[2]

Thus, homogeneity as a basis of grouping appears wide-spread. An elementary teacher in a school with one

---

[1] William G. Brink, "Secondary Education--Programs," in The Encyclopedia of Educational Research, 3rd ed., ed. by Chester W. Harris (New York: The Macmillan Company, 1960), p. 1267.

[2] J. Lloyd Trump and Delmas F. Miller, Secondary School Curriculum Improvement (Boston: Allyn and Bacon, Inc., 1968), p. 5.

teacher per grade level tends to organize his class in groups in order to give more specialized instruction to those children who learn at relatively the same rate, as exemplified by reading groups. Intra-classroom grouping even in larger schools with multiple, homogeneously grouped classes at each grade level is a demonstration of instructors' concern for teaching groups of children who are even more alike in rate of progress than the grouping policies of the school tend to indicate.

Grouping for instruction is an apparent attempt to recognize and utilize individual differences which exist among children.[3] The concept of individual differences appears to be operative at all chronological age levels. Usual administrative devices for grouping, however, often apply only to the selection of pupils. The assumption seems to be that teachers are equally effective regardless of their assignment to subject or group. Many authors suggest this area as a fruitful field for study.[4]

---

[3]Glen Heathers, "Grouping," in The Encyclopedia of Educational Research, 4th ed., ed. by Robert L. Ebel (London: The Macmillan Company, 1969), p. 564.

[4]A. S. Barr, "The Measurement and Prediction of Teacher Efficiency: A Summary of Investigations," Journal of Experimental Education, XVI (June, 1948), 206; A. S. Barr, et al., Wisconsin Studies of the Measurement and Prediction of Teacher Effectiveness (Madison, Wisconsin: Dembar Publications, Inc., 1961), p. 121; Roy M. Hall, "Staff--Selection and Appointment," in The Encyclopedia of Educational Research, 3rd ed., ed. by Chester W. Harris (New York: The Macmillan Company, 1960), p. 1377; and Harold E. Mitzel, "Teacher Effectiveness," in The Encyclopedia of Educational

Teacher assignment to homogeneously grouped classes often appears to be made on the basis of the general social acceptance of the slow-learning child. So that no teacher will be classified, by association, as "retarded," a school may employ a system of yearly rotation in teacher assignments. To one teacher may be assigned the slow learners for the first year, the average learners for the second year, and the gifted learners for the third year. Often a repetition of this three-year cycle occurs.

Such mechanical operations preclude any consideration of the instructor's effectiveness with certain groups, other than advantages occurring by chance alone. This three-year system of rotation may nullify any advantages that can accrue if individual differences are recognized among both the learner and the instructor corps.

Heretofore, most research based on product evaluation (evaluation of the student) has held as an assumption that a teacher, if he is "good" or if he is "bad" in one subject, grade, or level, is likewise "good" or "bad" in all subjects, grades, or levels. Some school systems even base ratings or rankings of teachers on the teachers' "overall" or "general effectiveness," ignoring completely the possibility of exceptional performance in one aspect of their instruction.

Research, 3rd ed., ed. by Chester W. Harris (New York: The Macmillan Company, 1960), p. 1485.

Davis and Nickerson propose that there is

> . . . no such thing as a good teacher as measured in universal terms. Teacher effectiveness can be evaluated only in its relationship to the contribution to the total staff effort . . . .
>
> . . . . Recruitment and selection of teachers can be accomplished effectively only after a thorough assessment of the unique teacher capability required to enhance the competence of the total staff . . . . . The legitimacy of the evaluation must rest upon . . . determining where the individual teacher's capacity best serves the total learning process of the school.[5]

Accordingly, some form of evaluation which will be an indicator of teacher effectiveness within the system should be devised to facilitate the placement of the instructor in his optimum position in the system.[6]

Therefore, a purpose of this study was to provide exploratory research to discover if a psychological principle of individual differences actually exists among teachers when "capacity to teach" is defined and when measures are employed to point out these differences. If capacities to teach were found to differ among the school staff members, a second purpose was to discover if a teacher's preference for an ability level of students would be an adequate indicator of the teaching position to which that instructor should be assigned.

---

[5]Donald E. Davis and Neal C. Nickerson, Critical Issues in School Personnel Administration (Chicago: Rand McNally and Company, 1968), pp. 1-2.

[6]Bruce J. Biddle and William J. Ellena, eds., Contemporary Research on Teacher Effectiveness (New York: Holt, Rinehart and Winston, 1964), p. 98.

Problem

This investigation was designed to test whether teacher preference for an indicated ability grouping reflects the achievement of that ability level of students of the instructor when compared with school norms.

Definitions

Teacher Effectiveness or "Capacity to Teach"

To describe an indicator for "a capacity to teach" a certain classification of pupils, an analyst would be required to investigate a multitude of ramifications attending the teacher's preparation, experiences, and personality. Research failed to reveal a valid and reliable instrument for measuring and evaluating these attributes as they relate to effective instruction. In the absence of such a device, a less rigorous, but nonetheless, beneficial system of investigation was employed.

Rather than calculating a teacher's latent potential, this inquiry was proposed as an investigation of an aspect of the teacher's product--the academic achievement of children assigned to him for instruction. In the process of teaching, the full potential of a teacher is brought to bear on a student with the intent of creating a product superior to that

of the untutored growth of the student.[7] Many authors recommend this product or pupil-change criterion.[8]

Teacher effectiveness, therefore, is defined as the result of a comparison of various achievement levels of the teacher's product with the corresponding levels of the product of the entire school. In this light, a teacher's effectiveness was considered as plus or minus when compared with the school norm, and no value judgement was made as to the magnitude of the differences. Comparisons of the degree of effectiveness between teachers were, also, of no concern.

Grouping

In any subject or grade level in which students are assigned to two or more teachers, some form of grouping must be employed. This may be accomplished by formal assignment or informal grouping, homogeneous or heterogeneous placement, or grouping into classes of various sizes. A number of other variables may be employed. Whatever the method, the students are organized into groups.

---

[7]Barr, "The Measurement and Prediction of Teacher Efficiency: A Summary of Investigations," p. 205.

[8]William H. Burton and Leo J. Brueckner, Supervision (New York: Appleton-Century-Crofts, 1955), p. 330; Davis and Nickerson, Critical Issues in School Personnel Administration, p. 65; Kathleen M. Evans, "A Critical Survey of Methods of Assessing Teaching Ability," British Journal of Educational Psychology, XXI (June, 1951), 89; Loren R. Tomlinson, "Recent Studies in the Evaluation of Teaching," Educational Research Bulletin, XXXIV (October, 1955), 1978; and James A. Van Zwoll, School Personnel Administration (New York: Appleton-Century-Crofts, 1964), p. 258.

Homogeneous Grouping

Homogeneous grouping can be based on numerous criteria wherein the students possessing the designated feature or degree of feature would be assigned to one group, and those not possessing the feature to another group. The criterion used in this study was the student's capacity to achieve. Three groups or ability levels are described (high, medium, low) by dividing the range of scores into thirds.

Capacity to Achieve

Capacity to achieve is defined as the ability to produce a certain level of achievement on standardized tests of subject matter. The index of this capacity is the score on a test of mental ability.

Achievement

Achievement refers to the level of performance in subject matter areas as measured by composite scores on standardized tests of subject matter.

Ability Level

Ability level reflects the homogeneous grouping of students by scores on mental ability tests. The term can describe five-point ranges of intelligence quotient used for statistical purposes, or it can describe the three homogeneous grouping levels (high, medium, and low) which were arbitrarily chosen because of their frequent use in schools employing homogeneous grouping.

Hypotheses

The first hypothesis stated in null form reads: When measured against the achievement of their pupils, all teachers are equally effective instructors throughout the range of pupil abilities. This hypothesis was to be rejected if the teacher norms exceeded and fell below the school norms. The alternate proposition would then be accepted: When measured against the achievement of the pupils, not all teachers are equally effective instructors throughout the range of pupil abilities. There are teachers whose instruction appears to produce more learning among those children who are working below grade level. The pupils of other teachers apparently progress more substantially if the children's capacity to achieve is in the normal range. Still other teachers are at the peak of their effectiveness if assigned children with superior capacities. Additionally, there may be those teachers who are able to produce exceptional achievement with all levels of students; and conversely, there may be those whose product, achievement-wise, is consistently lower than might be anticipated by a statement of the capacity to achieve of the children assigned to them.

A second null hypothesis was examined: The teacher's choice from among various possible ability level groupings does not correspond to the teacher's area of effectiveness as indicated under the first proposition. Statistical procedures were employed in the comparisons of teacher

preferences with norm outcomes to determine rejection. The alternate hypothesis states that the teacher's free choice of group level tends to be a useful predictor of effectiveness in the assignment of teachers.

Limitations

As this research was done in a single school system, the findings may not be generalized. However, the procedure might be applied in any school situation meeting the necessary criteria. This study is proposed as exploratory in nature in light of the limited number of teachers participating.

"Teacher effectiveness" is defined in terms of pupil achievement, only. It is generally accepted that achievement is not the only product which is desired from schools today. Pupil achievement should not be an exclusive factor when the productivity of teachers is examined.[9] This study is designed as one tool among many which might be employed to improve educational output.

There are many variables at work within the teachers, the students, and the classroom environments which can have effect on learning. This probe did not seek to identify and control these factors. Instead, the factors were allowed to influence the final product (the achievement of the students), which was the basis of measurement.

---

[9]William B. Ragan, Modern Elementary Curriculum (3rd ed.; New York: Holt, Rinehart and Winston, 1966), p. 453.

The measuring instruments involved have been standardized nationwide. An investigation was made of the validity and reliability of these measures. Data were drawn from tests published by a single firm in each instance--mental ability and achievement.

Achievement measures used in this work were composite scores of all the subject area sections. Individual subject matter areas were not an element of this paper. This exploratory probe was directed toward discovering whether differences existed in a teacher's general capacity to teach a variety of ability levels of students.

CHAPTER II

REVIEW OF THE LITERATURE

From Mitzel's description in The Encyclopedia of Educational Research, it appears that criteria for the judgment of teacher effectiveness have been proposed, rejected, and disputed in education for many years.[10] Different periods have seen one form rise in popularity only to become de-emphasized later.

Mitzel reports three classifications of criteria for evaluation of teaching: product, process, and presage. Product criteria are concerned with the effects of a teacher on the student. Operationally, interest is focused on the students' behavior as it progresses toward defined goals of the teaching situation. Definitions of goals and measurement of progress toward those goals can be very difficult with these criteria.

Process criteria seek to define those aspects of teacher and student behavior which seem to have a catalytic effect on product criteria. Conditions within the classroom or exhibited behavior are examples of this type.

---

[10] Mitzel, "Teacher Effectiveness," pp. 1482-85.

Presage or presumed criteria consist of variables which can possibly be assumed to have a relationship to the product. These factors bear no close relationship to the goals or procedures of education. Four common types of presage variables include teacher personality attributes, teacher training, teacher knowledge, and teacher status.

A review of research in annotated bibliographies confirms the use of all three of these classifications for evaluation.[11] However, not all criteria types are thought equally effective.

Many authors indicate that the product criteria would logically be the most important and useful. Mitzel states, "If certain definable and observable educational means are clearly better than others, then their effects should be discovered in measured educational ends."[12] "Product criteria depend for definition upon a set of goals toward which teaching is directed."[13] Ackerman says that the ultimate criteria is pupil change in behavior.[14]

---

[11]Simeon J. Domas and David V. Tiedeman, "Teacher Competence: An Annotated Bibliography," Journal of Experimental Education, XIX (December, 1950), 101-19; and William A. Watters, "Annotated Bibliography of Publications Related to Teacher Evaluation," Journal of Experimental Education, XXII (June, 1954), 351-69.

[12]Mitzel, "Teacher Effectiveness," p. 1484.

[13]Ibid., p. 1483.

[14]William I. Ackerman, "Teacher Competence and Pupil Change," The Harvard Educational Review, XXIV (Fall, 1964), 274.

Mitzel states, "Considering the theoretical importance of product criteria in the assessment of teacher effectiveness, it is surprising that so few studies have used some measure of student growth as the operational definition of teacher competence. Barr's summary of 138 studies published in 1948 lists only 19 . . . . In 1956 Mitzel and Gross found only 20 studies which had used a student-growth criterion . . . ."[15] Out of 1006 items in the Domas-Tiedeman bibliography, only 36 or 3.6 per cent looked at the "difficult approach" of pupil achievement.[16]

Possibly two of the earliest attempts at applying this form of criteria were executed by McCall and, later, Jenkins. Not only did these procedures assume that teacher effectiveness could be measured by the achievement of pupils, but they also converted these achievement scores into comparable figures by allowing for pupil capacity as related to achievement. "Achievement quotient" or "A.Q." was determined by:

1. determining the score in the academic area for each pupil
2. converting the score to an educational age
3. determining each pupil's mental age
4. finding each student's I.Q.

$$\left(\frac{MA}{CA} \times 100\right)$$

5. dividing educational age by mental age to get A.Q.

---

[15]Mitzel, "Teacher Effectiveness," p. 1483.

[16]Committee on the Criteria of Teacher Effectiveness, Report of the Committee, Review of Educational Research, XXII (June, 1952), 258.

6. estimating final mental age for the end of the teaching period
7. determining the final academic area score at the end of the teaching period
8. converting this score to a final educational age
9. dividing the final educational age by the estimated final mental age to get a final A.Q.
10. subtracting the mean initial A.Q. in the academic area from the mean final A.Q. in that area.

If the mean difference is 0, the teacher is average. If the difference is positive, the teacher is superior; and if a negative result is obtained, the teacher seems below average. In addition, a total teacher efficiency rating could be obtained by summing the A.Q. differences of all academic areas and dividing by the number of areas.[17]

Intelligence Quotient or Mental Age as an Indicator of Capacity to Learn

The use of mental age as an influence in capacity to learn is amply evidenced by McCall's "achievement quotient." Day's "teaching quotient" and Stephens' and Lichtenstein's "class and individual efficiency" are variations or extensions of McCall's procedure.[18] This concept can be substantiated

---

[17]William A. McCall, How to Measure in Education (New York: The Macmillan Company, 1922), pp. 149-68; and Albian U. Jenkins, "Measurement of Teaching Efficiency by Means of Standardized Tests," Department of Elementary Schools Principals Bulletin, VIII (1929), 373-82.

[18]McCall, How to Measure in Education, pp. 152-53; William A. McCall, Measurement of Teacher Merit (Raleigh, N.C.: Department of Public Instruction, 1952), pp. 19-20; L. C. Day, "The Teaching Quotient," The Elementary School Journal, XXXIII (April, 1933), 604-07; and J. M. Stephens

by evidence presented by Terman of a .725 correlation between mental age and quality of work in the first grade, and McCall's findings of a .78 correlation in the sixth grade.[19] Michael states that ". . . standardized measures of general or abstract intelligence . . . are extremely useful in furnishing an indication of what level of performance might be expected of students in school . . . ."[20]

Angoff proposes possibly the greatest difficulty in the comparison of mental ages to intelligence quotients. He states, "An I.Q. is meaningful only if there is an age for which the given mental age is average."[21] This, of course, has more implications at the age of fourteen or above as intelligence quotient performance tends to level off at this age. However, this is also a caution against comparisons of students with wide ranges of chronological ages.[22]

---

and Arthur Lichtenstein, "Factors Associated with Success in Teaching Grade Five Arithmetic," Journal of Educational Research, XL (May, 1947), 683-94.

[19]Lewis M. Terman, The Intelligence of School Children (Boston: Houghton Mifflin Company, 1919), p. 64; and McCall, How to Measure in Education, p. 21.

[20]William B. Michael, "Aptitudes," in The Encyclopedia of Educational Research, 3rd ed., ed. by Chester W. Harris (New York: The Macmillan Company, 1960), p. 59.

[21]William H. Angoff, "Measurement and Scaling," in The Encyclopedia of Educational Research, 3rd ed., ed. by Chester W. Harris (New York: The Macmillan Company, 1960), p. 815.

[22]Ibid.

Consistency of the Intelligence Quotient

It is now generally recognized that a child's measured intelligence quotient remains relatively consistent.[23] Burks summarizes several studies which indicate that the test-retest correlations generally were in excess of .80.[24] Even with such consistency, changes of ten points are not uncommon. Fifteen-point changes may appear once or twice in one hundred cases.[25] The interval of time between the pre-test and the re-test affects, to some extent, the correlations.[26] "A time interval of from five to nine years between two Stanford-Binet tests increases the variability in I.Q. rating to almost twice that found when the time interval is less than two years [depending on the age of those tested]."[27]

---

[23]Lewis M. Terman, The Measurement of Intelligence (Boston: Houghton Mifflin Company, 1916), p. 68; and Read D. Tuddenham, "Intelligence," in The Encyclopedia of Educational Research, 4th ed., ed. by Robert L. Ebel (London: The Macmillan Company, 1969), pp. 660-61.

[24]Barbara Stoddard Burks, "A Summary of Literature on the Determiners of the Intelligence Quotient and the Educational Quotient," Nature and Nurture, Twenty-Seventh Yearbook of the National Society for the Study of Education, Part II (Bloomington, Ill.: Public School Publishing Company, 1928), pp. 319-25.

[25]R. S. Woodworth, Heredity and Environment (New York: Social Science Research Council, 1941), p. 82.

[26]Anne Anastosi and John P. Foley, Jr., Differential Psychology (New York: The Macmillan Company, 1949), p. 294.

[27]Ralph R. Brown, "The Time Interval between Test and Re-Test in its Relation to the Consistency of the Intelligence Quotient," The Journal of Educational Psychology, XXIV (February, 1933), 94.

Several factors which influence intelligence quotient change can be identified.[28] However, three factors are of most interest in this research. Time interval has been mentioned above, and the consensus is that time periods up to three or four years do not invalidate the results of a mental abilities test.[29] The age and the environment of the subject are the other two important considerations. As the age increases, the predictive qualities of the scores increase. The time between test and re-test can also be increased. Preschool tests (especially below age four) have little or no value in predicting adolescent and adult intelligence quotients. However, tests in the primary grades show relatively stronger correlations.[30]

Concerning the effect of environment on intelligence quotients, conflicting conclusions have been drawn from essentially similar evidence. Both works by Rogers, et al., and Terman indicate that environmental changes make no

---

[28]Janet Matthew and Bertha Luckey, "Notes on Factors that May Alter the Intelligence Quotient in Successive Examinations," Nature and Nurture, Twenty-Seventh Yearbook of the National Society for the Study of Education, Part I (Bloomington, Ill.: Public School Publishing Company, 1928), p. 412; and Terman, The Intelligence of School Children, p. 9.

[29]Tuddenham, "Intelligence," p. 661.

[30]Anastosi and Foley, Differential Psychology, pp. 293-96; and Martin J. Nelson, "Intelligence and Special Aptitude Tests," in The Encyclopedia of Educational Research, 4th ed., ed. by Robert L. Ebel (London: The Macmillan Company, 1969), p. 669.

significant changes in intelligence quotients.[31] In a probe concerning the relative merits of environment versus heredity, Burks reports that measurable (± one standard deviation) environmental changes affect intelligence quotients no more than six to nine points. The maximum contribution of environment to measured intelligence is apparently ± twenty points; however, these extremes might occur only once or twice in one thousand cases.[32]

Studies of twins and foster children in various environments have given comparative measures in the controversy concerning the role of environment in measured intelligence. Anastosi and Foley report several major studies of this type and interpret the results of these works as indicating no significant relationship between environmental changes and intelligence quotient.[33] Woodworth also reports the major works in this area. His conclusion holds that

---

[31]Agnes L. Rogers, Dorothy Durling, and Katharine McBride, "The Effect on the Intelligence Quotient of Change from a Poor to a Good Environment," Nature and Nurture, Twenty-Seventh Yearbook of the National Society for the Study of Education, Part I (Bloomington, Ill.: Public School Publishing Company, 1928), p. 330; and Terman, The Intelligence of School Children, p. 14.

[32]Barbara Stoddard Burks, "The Relative Influence of Nature and Nurture upon Mental Development: A Comparative Study of Foster Parent-Foster Child Resemblance and True Parent-True Child Resemblance," Nature and Nurture, Twenty-Seventh Yearbook of the National Society for the Study of Education, Part I (Bloomington, Ill.: Public School Publishing Company, 1928), pp. 308-09.

[33]Anastosi and Foley, Differential Psychology, pp. 342-47.

there have been consistent, though not necessarily significant, increases in intelligence quotients for those subjects in "better" environments. However, he stresses the fact that intelligence quotients are relatively stable in spite of mean gains up to ten points. A major conclusion is drawn that large differences between twins can inevitably be correlated to great contrasts in educational advantages.[34]

Relevant Assumptions of Past Studies

The first hypothesis of this investigation is basically an attack on a wide-spread assumption held in most past studies concerning teacher effectiveness. This assumption of the universality of teacher effectiveness seems to violate the concepts of individual differences. Few attempts have been made to differentiate individual strengths and weaknesses in teachers. Rather, a single indicator of effectiveness has been sought. Emphasis has been on finding the "best" teacher or a ranking of personnel.

Several other implicit assumptions are held in much of the earlier research. Most of these assumptions are based in the above universal effectiveness assumption. It has been held that a teacher teaches all subjects with equal effectiveness. An instructor is equally effective with all ability levels. "Good teaching" is irrevocably possessed by

---

[34]Woodworth, Heredity and Environment, pp. 29-30, 48-55, and 82-84.

a teacher and is in no way situational in nature. Since "good teaching" is of universal nature and a given measure of it is permanently instilled in each instructor, teachers can be compared in a completely reliable and valid manner. The teacher has been held responsible for the changes in the child's measured behavior. One or more of these assumptions can be found, for example, in four of the major research efforts: The Wisconsin Studies,[35] the McCall procedure,[36] the probe by Seyfert and Tyndal,[37] and the Bolton study.[38] These are not the only assumptions made by all investigators, but they are important as a basis for the problem in this work.

Individual Differences in Teachers

Although no research has been found which deals with teacher effectiveness variation between ability levels, two sources seem to indicate the concept of individual differences in teachers with respect to subject matter. Lancelot, et al., indicate that teachers vary as to which courses they

---

[35]Barr, Wisconsin Studies of the Measurement and Prediction of Teacher Effectiveness, passim.

[36]McCall, How to Measure in Education, passim.

[37]Warren C. Seyfert and Balfour S. Tyndal, "An Evaluation of Differences in Teaching Ability," Journal of Educational Research, XXVIII (September, 1934), 10-15.

[38]Floyd B. Bolton, "Evaluating Teaching Effectiveness Through the Use of Scores on Achievement Tests," Journal of Educational Research, XXXVIII (May, 1945), 691-96.

teach best.[39] Jones reveals that instructors teach specific areas of a given subject better than other areas.[40]

Teacher Prediction of His Effective Area

With respect to the idea that a teacher is able to predict or choose the area in which his best work is done, Watters reports an investigation that confidential teacher self-ratings held the best correlation to teaching effectiveness.[41] Thelen, in working with his "teachability groups," reports that teacher choice of the type of student he had in class accomplished several seemingly desirable goals. Among these were: the teacher was more satisfied, there were fewer discipline problems, and the class received higher grades. As no analysis was done to investigate improvement in achievement, no conclusions in this area were drawn.[42] However, psychologists such as Combs might argue that the learning atmosphere could have great effect on the amount of achievement.[43] On the basis of these works, a researcher might find a basis for testing whether a teacher could predict his area of effectiveness.

---

[39]William H. Lancelot, et al., The Measurement of Teaching Efficiency (New York: The Macmillan Company, 1935), pp. 1-61.

[40]E. S. Jones, "Suggestion for Teacher Measurement," School and Society, VI (1917), 321.

[41]Watters, "Annotated Bibliography of Publications Related to Teacher Evaluation," pp. 351-69.

[42]Herbert A. Thelen, "Grouping for Teachability," Theory into Practice, II (1963), 81-89.

[43]Arthur W. Combs, The Professional Education of Teachers (Boston: Allyn and Bacon, Inc., 1965), pp. 98-111.

CHAPTER III

PROCEDURE

Population Description

The total eligible enrollment of students in the second and third grades during the school years from 1966 through 1970 in the Duncan, Oklahoma, Public School System comprised the basis for product measurement. The teachers who had taught all of these years exclusively in the second or in the third grade became the units of analysis of this investigation.

Duncan is a southern Oklahoma community of approximately 20,000 population. Its industries center about oil production. There seems to be the entire range of socio-economic levels present in this city. Major racial groups are Caucasians, Negroes, and Indians. Other races are also represented.

The seven elementary schools do not differ radically in size. The smallest contains no fewer than one teacher per grade, while the largest has two and one-half teachers per level. The classes are grouped heterogeneously. departmentalization exists only in the fifth and sixth grades.

Each teacher, to be eligible for use in this study, must have taught either in the second or in the third grade for all of the four years under investigation. The four-year period was arbitrarily chosen because a longer period would have drastically reduced the number of eligible teachers. The second and third grades were chosen as the levels for the study because of the greater number of eligible teachers in these grades.

There was an average of twenty-eight teaching positions in these two grades each year. Because of movement to and from the city and grade placement changes within the system, a total of forty-three teachers taught in these grades at some time within the period. Even though several teachers had been employed for the full period, because of movements to positions in different grade levels, many were lost to the study. A total of eleven teachers was found to have met the time and level criteria. Five of these taught the second grade, and six taught the third.

The data of teachers who were found not to meet the two criteria were excluded from the computations. Their students' scores, nevertheless, were used in the calculation of the school norm.

The total population of students in the second and the third grade within the four-year period was used as a parameter. However, certain students' scores were not used. Many moved into or out of town or from teacher to teacher so

that a total of one hundred days within a year in a single teacher's room was not achieved. Some students were not present on the days of testing for intelligence quotient or achievement. A few pupils did not attain a valid score on one or both tests. Finally, a few children were disqualified because their age exceeded, by twelve months, the "normal" age of that grade level. Of the students in the second grade, approximately 32 per cent were disqualified on the above grounds, leaving a total of 1063 eligible. About 30 per cent of the third grade was found insufficient, yielding a remainder of 1091 useful students. A tabulation of the total number of eligible students in each teacher's room can be found in Appendix III, page 75ff.

Outline of the Procedure

The procedure for discovering a teacher's capacity to teach at different ability levels of a specific category was to:

1. determine a school norm of achievement for children of a specified grade level based on the mental ability of the children
2. discover the norm of achievement for the individual teacher's former pupils, using identical measures as were employed in establishing the school norm
3. compare the teacher's norm(s) with those of the school.

The procedure for determining the validity of the teacher's opinion as a predictor of effectiveness was to:

1. present the teacher with a questionnaire concerning his choice of ability level grouping.

2. compare the results of the questionnaire with the results of the first part of the study for each teacher and determine the statistical significance and relationship of these two measures.

Design

A form of ex post facto research design was employed. This method, employing recapitulation, offers a measure of protection against manipulation of the results. If the study were a projection, rather than ex post facto, a teacher might have been able to direct his instruction to the area which he chose on the questionnaire.

The design consisted of parametric data (the total student body of the second and the third grades of the Duncan Public Schools for four years) from which samples (each teacher's group of students for that four-year period) were drawn. The statistics and results of this study are therefore not to be generalized outside of the limits of the populations as described. The procedure, however, may have possibilities for use in other situations.

This design did not employ a control group procedure. Instead, this study used the group as its own control. This internal control aspect of comparing samples to the total population lends strength to the over-all design in that the control group's inherent lack of identity with the experimental group did not apply.

Method of Collecting Data

Two scores for each student are needed for the performance of this procedure: an intelligence quotient score from a standardized test of mental ability, and a composite achievement score from a standardized test of achievement taken near the end of the year being studied. If central records of the scores are kept in a school system, the data would be readily available. However, the Duncan school system was in the process of compiling such central files at the time of the study. They had finished the files of achievement scores, but had not yet completed the mental abilities scores. It was necessary to examine the permanent record folder of every child at each school as well as the central "dead files" or records of those who had left the system. To insure completeness, the names of the students found in the files were checked against the complete central record of achievement test scores. Those students who were in the system but not on the achievement test score list were, of course, not eligible for the study, as that score is necessary to the performance of the procedure.

Since tests were done at the same time each year, interpolation to equalize data was not necessary. These tests were given in Duncan on approximately April first.

As intelligence quotients appear quite stable, no significance was assigned to the date of the child's test of mental ability. Although radical changes in various factors

can influence intelligence quotients, there appeared to be no evidence of such forces having affected the children in Duncan. In the Duncan elementary school system, mental ability tests were given in the first, third, and fifth grades. This study used the third grade scores as ability indicators for both second and third graders. All subjects in the second grade study were at least in the third grade at the time of the research, as the investigation did not use data from the 1970-71 school year.

To improve the validity of the data by increasing the number of students involved, a survey of achievement marks within the chosen grades for a four-year period was used. The 1966-67, 1967-68, 1968-69, and 1969-70 school years were involved. This device--an arbitrary four-year coverage--served as a partial safeguard against errors in trends of achievement of a particular group which might have reflected a poor or highly favorable grouping pattern of a single year's duration.

The final piece of necessary data--the response of each teacher indicating his choice of ability levels--was obtained through individual interviews conducted by the investigator. Each teacher was given a single-item, oral questionnaire concerning his preference for a level of students. His response to this question, as well as other comments which the interviewer felt were significant, were recorded to provide nominal data for the final comparison of

the validity of teacher prediction of effectiveness. The interview and questionnaire are discussed in more detail in another section to follow.[44]

Test Instruments

The Otis-Lennon Mental Ability Test was used to provide an intelligence quotient for the students. The Elementary I level of Form J of this teacher-administered test is described as a measure of the general intellective ability factor and does not measure innate mental capacity. The machine-scored, timed test was normed on 200,000, K-12 grade pupils in the fall and winter of the 1966-67 school year. This sample included .4 per cent of the total population.

Reliability was determined by the use of three tests. Coefficients for the second grade are .88 for the Kuder-Richardson test, .89 for the split-half correlation, and .85 for the alternate forms procedure. Corresponding correlations for the third grade were .91, .92, and .89. Data is currently being tested to show validity of the test.[45]

A composite achievement score for each student was obtained from Forms W and X of the California Achievement Tests Complete Battery. This upper primary test, composed of tests in reading, arithmetic, and language was re-normed

---

[44] *Vide infra*, p. 33.

[45] Arthur S. Otis and Roger T. Lennon, *Otis-Lennon Mental Ability Test, Form J* (New York: Harcourt, Brace and World, Inc., 1967), pp. 4-21, *passim*.

in 1963 on 15,351 1-12 grade students. A total of 1884 subjects was retested on both this instrument and the California Achievement Tests short form.

Reliability on the Kuder-Richardson Formula 21 was found to have a coefficient of .98. Content validity is based on refinement of the instrument by phi coefficient discrimination analysis since 1937. Construct validity showed a high correlation coefficient of .43.[46]

The Duncan school system used both of these tests exclusively during the period under investigation. However, neither test was used in the 1965-66 school year (the year preceding this study), and the California Achievement Test was not used in the 1970-71 school year. These facts influenced the choice of a four-year period rather than a longer amount of time as a basis for norms in this research.

School Norm

To determine a school norm for achievement, two kinds of data were required for each pupil studied. The intelligence quotient of the child, obtained from the Otis-Lennon Mental Abilities Test, was used as an indication of his capacity to achieve, and the composite grade equivalent score the child made on a standardized test of academic achievement (The California Achievement Test Battery) served as a measure

---

[46]Ernest W. Tiegs and Willis W. Clark, devisors, California Achievement Tests Complete Battery and Separate Reading, Arithmetic, and Language Tests, 1957 ed. (Monterey, Calif.: California Test Bureau, 1957), pp. 5-36, passim.

of the school's product. The composite achievement score was used rather than individual subject area scores; thus, differences in teacher ability by subject were not examined.

Basic categories of intelligence quotients were arbitrarily fixed at five-point intervals, for example, intelligence quotient category 95-99. All students were assigned to their indicated interval group. Only one grade level was studied for each comparison of norms. Students whose ages exceeded, by over twelve months, the normal age range for the grade under study, were not included in the norm computations.

In processing the data, the achievement grade equivalent scores for all pupils in each intelligence quotient category were compiled and a mean determined. The achievement norm for the school was arbitrarily established as this mean. In like manner, all intelligence quotient categories were figured.

Teacher's Norm

The norm of achievement for a teacher's pupils was discovered by a process identical to that used to find the norm of the school. Only those cases which have been pupils of the eligible teacher applied.

Norm Comparison Procedure

This portion of the study was designed to use as little statistical procedure as possible. Other than score

manipulation and means computation, no statistical tests of significance were employed. Two reasons could be proposed for this procedure. First, since effectiveness was defined as plus or minus the school mean, the degree of plus or minus was not to be evaluated on its degree of significance. Interest was focused on the levels in which a teacher's norm exceeded the school norm on the premise that it seemed rational to place a teacher in the indicated area of his best work, ignoring the degree of deviation from the mean of the "best" area. Second, it was thought that the simplicity of the design may lead to ease of application in other situations.

The procedure for the comparison of the teacher's norms to the school norms was visual in nature. The school norms for all ability levels were determined and listed in one table for each grade. Each teacher also had a table listing his means within each intelligence quotient level. The teacher's table and the school's table for that grade were then compared level by level. If the teacher's mean exceeded the school mean in any level, a check was placed at that level on the teacher's table. No mark was made when the means were equal or when the teacher's mean fell below the school mean. The degree of difference between the teacher's norms and the school norms was of no interest in this study.

To find the points which divide the grades into thirds, the total number of students was divided by three to find

the number in each third. Starting at either extreme of intelligence quotient categories on the school norm table, the number of students in each five-point level was accumulated until the one-third point was found. The procedure was then repeated for the other extreme. Those levels were indicated on the tables as the dividing points of the thirds. Unless a dividing point fell on the break between five-point levels, the level in which it fell was considered in the determination of effectiveness for both thirds involved.

Only those areas on the norm tables where the teacher norms exceeded the school norms were considered. To have achieved effectiveness for the entire third of the class, the teacher norms must have exceeded the school norm in at least 50 per cent of the intelligence quotient categories in which he had students within that third of the class. For instance, if the instructor had students in six intelligence quotient levels in the lower third of the class, the instructor is said to be effective with that third of the class when the teacher norm exceeded the school norm in three or more of these levels.

An "effective level" or "effective third" on the table had to be present to correspond to the response on the questionnaire to be considered a match. Each teacher had only a single chance for a match. A teacher could match his choice in one third and also be shown effective in another area in

the table, but a match did not occur unless the level selected on the questionnaire was also indicated as effective on the table.

A visual test was applied to determine whether rejection of the first null hypothesis was possible. For acceptance of the null hypothesis, each teacher should have been either effective or not effective across all ability levels. If any teacher's norms exceeded and fell below the school norms, the alternate hypothesis should have been accepted.

Interview and Questionnaire

Each individual teacher was privately interviewed by the investigator. The aim of the interview was to attempt to increase the reliability of the questionnaire by informing the teachers of the idea and mechanics behind the study. However, only a certain amount of information was discussed before the teacher was asked to complete the questionnaire. This information consisted of a basic statement concerning the purpose of the study, assurances of anonymity, and assurances that the results of this study would not adversely affect the teacher in any way. The purpose of this personalized interview was to reduce the fear of recrimination which might have erroneously occurred to the teacher.[47]

---

[47]Davis and Nickerson, Critical Issues in School Personnel Administration, p. 66.

In addition, emphasis was placed on the fact that stigmas have, in the past, been abundant in the area of teacher placement. Attempts were made to point out that the teacher of the low mental ability groups is as important as any other teacher. Also, mention was made of the past emphasis which was placed on teaching the upper ability level group. A complete list of the points covered can be found in Appendix I, pages 71-72.

The questionnaire consisted of a single, multiple-choice item with five possible answers:

If I had my choice, I would prefer to teach the following level:

1. the lower third of the second (or the third) grade based on mental ability
2. the middle third of the second (or the third) grade based on mental ability
3. the upper third of the second (or the third) grade based on mental ability
4. a class comprised of all of the above three levels
5. none of the above four choices for any reason.

The question was read to teachers and any explanations which were needed were provided. Two examples were used for the fifth choice: the teacher who is teaching one grade and wishes to teach another, or a teacher who does not really wish to teach. No explanation of the instructor's response was required.

Following the administration of the questionnaire, each teacher was asked to reveal nothing concerning the interview and the questionnaire for that period of time the investigator estimated would be required for the completion of the

data gathering. This was done in an attempt to allow each teacher an equal amount of time for thought concerning his response.

Comments which the teachers made concerning the study, the philosophy behind the study, or their choices were noted on the same sheet which held the marked questionnaire. These comments were recorded in the event that they might prove worthwhile in the analysis of the data. The original of this sheet was given to the teacher and a carbon copy was retained by the interviewer.

Statistical Comparisons of Indicators of Teacher Effectiveness

A contingency coefficient "C" was employed to define the degree of relationship between teacher's choice and the norm indicators of teacher efficiency. This nominal statistic, based on chi square, can be used with no assumptions concerning the population distribution, normality, or scaling.

The data first had to be fit into a chi square matrix. The five-by-five matrix was chosen based on the number of choices on the questionnaire. The five columns represented those choices and the five rows depicted the corresponding outcomes from the norm tables as in Figure 1.

As defined earlier, it was possible for a teacher to be shown as effective in more than one level. Even though a teacher might have been effective with two levels, there

could have been factors which influenced the teacher's choice of one of these groups. These factors might not have been reflected in the achievement scores to such a degree as to cause the other level to seem non-effective. As the teacher, nevertheless, experienced these factors, his choice must be assumed to have indicated his best area if the table corroborates that area as effective.

Choice on the Questionnaire

Effective Level from the Mean Tables	Upper Third	Middle Third	Lower Third	All Levels	None of These
Upper Third					
Middle Third					
Lower Third					
All Levels					
None of These					

FIGURE 1

MATRIX FOR COMPARISON OF INDICATORS

Thus a procedure was defined to fit the data into the matrix. For each entry into the matrix, the teacher's choice determined the column to be used. If the norm table showed that the teacher was effective in the chosen level a "1" was placed in the corresponding row. Even if the teacher was shown effective in other levels, these were disregarded.

For example, if a teacher answered the questionnaire with "Middle Third" and was shown effective in both "Upper Third" and "Middle Third," the investigator worked within column "Middle Third" because of the questionnaire response. He placed a "1" in the box in row "Middle Third" because of the match on the questionnaire and the norm tables. The fact that the norm table also indicated "Upper Third" was disregarded.

However, if the teacher's choice did not match the norm table results, another procedure was used. Again, the teacher's choice on the questionnaire determined the column to be used. If the teacher was shown effective in any one level, a "1" was placed in that numbered row within the predetermined column. If the instructor appeared effective in two areas and still made no match, ".5" would have been placed in the rows corresponding to both of these levels. For example, if a teacher chose "Lower Third" but was shown effective in "Upper Third" and "Middle Third," the marks would have been made within column "Lower Third." Since two effective levels were indicated, ".5" would have been written in both row "Upper Third" and row "Middle Third."

This procedure was necessary so that the total of the rows and the total of the columns would each equal eleven--the number of teachers involved. Since it was possible to be effective in more than one level, this form of weighting was employed to provide the necessary row total.

A theoretical or "expected" five-by-five matrix was then calculated based on the column and row totals of the "observed" matrix. Each square was figured using the formula

$$\frac{\text{row}_x \text{ total X column}_y \text{ total}}{N}$$

where x was the row number which possessed the given box and y was the column number which held that same box. N was the number of teachers or 11. The degrees of freedom equalled

(number of rows - 1) X (number of columns - 1) or 16

Chi square was then found by calculating

$$x^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

For the five-by-five tables, one calculation for each pair of squares or twenty-five computations were summed. A test of the significance of this chi square also provided the test of the significance of the contingency coefficient. The .05 level was chosen as the confidence level for correlation.

To calculate the contingency coefficient, "C," the following formula was used:

$$C = \sqrt{\frac{x^2}{N + x^2}}$$

where $_{X}2$ equalled the chi square figured above and N was the number of teachers involved. Unlike many correlations, the coefficient of contingency can only approach 1 as an upper limit. This limit is based on the number of categories or rows and columns in the matrix. In this five-by-five matrix, k equalled the smaller number of either rows or columns--"5." The formula used for finding the upper limit was

$$\sqrt{\frac{(k - 1)}{k}} = \sqrt{\frac{5 - 1}{5}} = .894$$

Based on this five-by-five table, the coefficient could not have exceeded .894. The resultant coefficient indicated the degree of association between teacher's choice and the outcomes of the norm comparison procedure. The significance of this coefficient was found by determining the significance of the chi square.

Summary

Students in the second and the third grades in the Duncan, Oklahoma, Public Schools over the school years, 1966-70, were placed in five-point interval categories of intelligence quotient for each grade and the mean of standardized achievement test composite scores was determined for each intelligence quotient level to provide a school norm for each level. All teachers who had taught that given grade for those four years had their students drawn from that total group and the same procedure performed with the

test results. The norms for the teachers were then visually compared with the school norms to determine if individual differences existed in the teachers' capacity to teach when measured against the achievement of their pupils.

Teachers were given a questionnaire to determine their desire to teach a certain general ability level. A contingency coefficient, "C," based on the chi square statistic was used to determine whether a teacher's choice of ability levels appears to be a predictor of effectiveness when compared with the measurement of his product.

CHAPTER IV

FINDINGS

Development of the Means

To implement the examination of the hypotheses, norms were computed and tabled in a manner facilitating their comparison. The procedure was executed separately for the second grade and the third grade students. All students who were in the second or the third grade during the period under study and who met the criteria of having both necessary test scores and of being in the teacher's room for at least 100 days were included in the computations.

The students in each grade were self-selected into fifteen five-point intelligence quotient levels on the basis of their mental abilities test scores. The pupils' achievement test grade equivalent scores were then summed in each ability level and an average score for each level was determined by dividing the total of the achievement scores by the total number of students in that level. The results of these computations can be found in the tables in Appendix II, page 73ff.

The students of the eleven eligible teachers were then extracted from the total group. An identical procedure of

categorization into ability levels and computation of means was accomplished for each teacher. The results of these operations are reported in the tables of Appendix III, page 75ff.

Comparison of the Means

Each teacher's norms were then compared to the school norms for the corresponding grade. If the teacher's norm exceeded the school norm in an ability level, an "x" was placed by that level on the teacher's table in Appendix III.

Returning to the first hypothesis, if the null hypothesis that teachers are equally effective throughout the range of mental abilities of their students would have been true, a teacher would have had either "x's" in all levels or no marks in all levels on the tables in Appendix III. They would have been either effective or not effective with all ability levels of students. With only the exception of Teacher "H," all teachers had means both above and below the school norms. Based on this fact, the first null hypothesis was rejected. There appeared to be evidence that these teachers instructed students of varying abilities with dissimilar effectiveness when measured by the student achievement.

In an attempt to examine this variation in effectiveness further, a more stringent visual test was performed. For this additional computation, a standard deviation was calculated using the formula

$$sd = \sqrt{\frac{\sum x^2 - N\bar{x}^2}{N - 1}}$$

where $\sum x^2$ is the summation of each score in the ability level and $N\bar{x}^2$ is the number of students in the ability level times the square of the mean score of that level.

The results were recorded on the tables in Appendix II. Table 1, page 44, was drawn up illustrating the upper and lower limits of the school norms in each ability level when based on $\pm$ one standard deviation from the mean. The teacher's norms tables were then compared with this school mean $\pm$ deviation table. An asterisk was placed in the appropriately labeled column on the teacher's norms in Appendix III when a teacher's mean exceeded a plus deviation or fell below a minus deviation. It was found that four teachers exceeded the plus deviations a total of thirteen times and that three teachers fell below the minus deviation on five occasions. This supplemental, more stringent test added credence to the originally proposed procedure, confirming the evidence of individual differences within teachers. This latter test was used only to accentuate the variation in ability and bore no relationship to the determination of effectiveness as it was defined for the remainder of the investigation.

Determination of Effectiveness in Each Third

The table of school means for each grade was divided into thirds. In the second grade, one third of the total

TABLE 1

SCHOOL NORMS $\pm$1 STANDARD DEVIATION

Intelligence Quotient Level	Second Grade		Third Grade	
	Lower Limit	Upper Limit	Lower Limit	Upper Limit
Above 134	3.893	4.983	4.751	4.971
130-134	3.786	4.574	4.526	5.106
125-129	3.993	4.403	4.605	4.873
120-124	3.880	4.476	4.465	4.917
115-119	3.660	4.356	4.309	4.787
110-114	3.487	4.259	4.246	4.772
105-109	3.492	4.076	4.288	4.604
100-104	3.402	3.876	4.064	4.664
95-99	3.158	3.788	3.940	4.512
90-94	3.063	3.733	4.063	4.219
85-89	3.068	3.542	3.525	4.321
80-84	2.606	3.478	3.204	4.073
75-79	2.742	3.412	3.422	4.220
70-74	2.179	3.301	2.726	3.950
Below 70	2.315	3.651	2.606	4.460

students was found to equal 355 pupils. By summing the students in each ability level proceeding from the lowest level, this one-third point was found to fall in the 95 to 99 intelligence quotient level. Summing from the highest level, the lower limit of the upper third was located within the 110 to 114 ability level. Since the level in which a one-third point falls was to be counted in both thirds, it was found that the upper third included six five-point levels, the middle contained four levels, and the lower third encompassed seven ability levels.

Likewise, the third grade population was divided into thirds based on the figure of 364 students per third. The upper third extended downward into the 110 to 114 level, including six ability categories. The lower third exactly ranged up to, but not including, any of the 100 to 104 group, thus comprising seven levels. The middle third, therefore, consisted of three levels. Arrows were inserted in the tables in Appendices II and III to indicate the levels possessing the points of division.

For a teacher to be considered effective in any third, the "x's" in a third must have equalled at least 50 per cent of the total number of ability levels in which he had students in that third. For example, if a teacher in the second grade had students in all of the seven levels of the lower third, he must have had at least four "x's" on levels in that third to be described as effective. However, if that

same teacher had students in only six of those seven ability groups, his level of significance would have been reduced to three "x" marks.

On the basis of this procedure, Table 2 illustrates the effective thirds for each teacher. This table formed the basis for comparison of indicators of effectiveness.

TABLE 2

TEACHER EFFECTIVENESS, IN THIRDS, FROM MEAN TABLES

Teacher	Grade	Upper Third	Middle Third	Lower Third
A	2	x*		x
B	2			
C	2	x	x	x
D	2	x	x	x
E	2			
F	3	x		
G	3	x	x	x
H	3	x	x	x
I	3	x	x	x
J	3	x		
K	3			

*Effective in this third.

Results of the Questionnaire

The individual interviews with teachers were completed in a two-day period. All teachers were most co-operative and seemed intrigued with the idea behind the project. No teacher was especially interested in the procedure and statistics.

A check sheet was used in each interview to insure that certain points were mentioned before the questionnaire was administered. Those points are listed in Appendix I.

When the questionnaire was presented orally, the teachers seemed to know immediately what their choice would be. No teacher needed more than a few seconds thought before answering. Although no justification for a response was solicited, several teachers had comments on their choices. These were noted on the questionnaire sheet in case they proved worthwhile in the analysis of the data. The teacher was given the original of this sheet and a carbon copy was retained by the investigator.

Table 3 indicates the choices of the teachers from the five-item questionnaire. The totals at the bottom of each column show that all categories except the last ("None of These") were chosen by teachers, although not equally.

Comparison of Indicators of Effectiveness

To facilitate an analysis of the relationship between the teacher's choice and the results on his mean comparison with school norms, a five-by-five matrix was used. The five categories on the questionnaire were listed across the top

TABLE 3

TEACHERS' CHOICES ON THE QUESTIONNAIRE

Teacher	Grade	Upper Third	Middle Third	Lower Third	All Levels	None of These
A	2	x				
B	2		x			
C	2	x				
D	2	x				
E	2				x	
F	3	x				
G	3		x			
H	3		x			
I	3			x		
J	3	x				
K	3				x	
Totals		5	3	1	2	0

and the same five categories were used down the side as the levels of effectiveness from the mean tables, as in Figure 2. The column totals and the row totals should sum to eleven.

Choice on the Questionnaire

Effective Level from the Mean Tables	Upper Third	Middle Third	Lower Third	All Levels	None of These	
Upper Third	1, 1, 1, 1, 1 (5)	(0)	(0)	(0)	(0)	5
Middle Third	(0)	1, 1 (2)	(0)	(0)	(0)	2
Lower Third	(0)	(0)	1 (1)	(0)	(0)	1
All Levels	(0)	(0)	(0)	(0)	(0)	0
None of These	(0)	1 (1)	(0)	1, 1 (2)	(0)	3
	5	3	1	2	0	11

FIGURE 2

"OBSERVED" MATRIX

Tables 2 and 3 held the required data to complete the matrix. As a review of the procedure for entering the matrix, a teacher's choice on the questionnaire defined the column in which a mark would be placed. If his mean table effectiveness agreed with his choice, a "1" was placed in the corresponding row within that column. All other levels of effectiveness from the means tables were disregarded. However, if the instructor's mean table did not match his choice, a "1" was placed in the appropriate effective row within the

column defined by his choice. This procedure was used when only a single level of effectiveness or no effectiveness was achieved. When a match did not occur and two levels of effectiveness were present, a ".5" was placed within the column in both effective rows.

To illustrate, observe two examples--Teachers A and B. Table 3 showed that Teacher A chose the "Upper Third." On the matrix in Figure 2, this choice confined all marks to column "Upper Third." Table 2 indicated that this instructor was shown effective in both the upper and the lower thirds. A match was therefore found between the two indicators in the upper third. A "1" was placed within column "Upper Third" in row "Upper Third." The effectiveness of the teacher in the lower third was disregarded.

Teacher B chose the "Middle Third" on the questionnaire, confining all marks to column "Middle Third." Table 2 illustrates that this teacher was not shown effective in any level. Therefore, a "1" was placed within column "Middle Third" in row "None of These." Figure 2 shows the final results of the tally for all teachers. This became the "observed" matrix for computation.

An "expected" matrix was computed based on the column and row totals in Figure 3. For each square, the corresponding column total was multiplied by the corresponding row total and the resulting figure divided by N (11). A chi square test was then able to be performed between the "expected" and "observed" matrices.

Choice on the Questionnaire

Effective Level from the Mean Tables	Upper Third	Middle Third	Lower Third	All Levels	None of These	
Upper Third	2.27	1.36	.46	.91	.00	5
Middle Third	.91	.55	.18	.36	.00	2
Lower Third	.46	.27	.09	.18	.00	1
All Levels	.00	.00	.00	.00	.00	0
None of These	1.36	.82	.27	.55	.00	3
	5	3	1	2	0	11

FIGURE 3

"EXPECTED" MATRIX

A chi square was performed using the formula

$$x^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

The frequency in the square of the "expected" matrix was subtracted from the number in the corresponding square of the "observed" matrix. That result was squared and divided by the "expected" frequency. Chi square was found by summing the results of the computations for all twenty-five squares of the matrix as in Table 4.

TABLE 4

COMPUTATION OF THE CHI SQUARE

Observed (O)	5.00	0.00	0.00	0.00	0.00	0.00	2.00	0.00	0.00	1.00	0.00	0.00	1.00
Expected (E)	2.27	.91	.46	.00	1.36	1.36	.55	.27	.00	.82	.46	.18	.09
$\lvert O-E \rvert$	2.73	.91	.46	.00	1.36	1.36	1.43	.27	.00	.18	.46	.18	.91
$(O-E)^2$	7.45	.83	.21	.00	1.85	1.85	2.10	.07	.00	.03	.21	.03	.83
$\frac{(O-E)^2}{E}$	3.28	.91	.46	.00	1.36	1.36	3.82	.27	.00	.04	.46	.18	9.22

TABLE 4--Continued

Observed (O)	0.00	0.00	0.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00	0.00	0.00
Expected (E)	.00	.27	.91	.36	.18	.00	.55	.00	.00	.00	.00	.00
$\|O-E\|$	.00	.27	.91	.36	.18	.00	1.45	.00	.00	.00	.00	.00
$(O-E)^2$	.00	.07	.83	.13	.03	.00	2.10	.00	.00	.00	.00	.00
$\frac{(O-E)^2}{E}$	.00	.27	.91	.36	.18	.00	3.82	.00	.00	.00	.00	.00

$$x^2 = \sum \frac{(O-E)^2}{E} = 26.90$$

The resultant chi square was found to equal 26.90. The degrees of freedom was found by

(Number of rows - 1)(Number of columns - 1)

or

$$(5 - 1)(5 - 1) = 16$$

The result of the chi square computation was used as a basis for finding the contingency coefficient "C." The formula

$$C = \sqrt{\frac{x^2}{N + x^2}}$$

was used by inserting the correct figures. Thus the coefficient was found to be

$$C = \sqrt{\frac{26.90}{11 + 26.90}} = .843$$

As calculated earlier, the maximum upper limit for this correlation coefficient was found to be .894. The test of significance for this correlation was the test of significance for the chi square. With a df of 16, the chi square of 26.90 was significant beyond the .05 level of confidence. Thus a significant relationship was found between teacher choice of ability level and the results of the norm comparisons for effectiveness.

Summary of the Findings

Means were determined for both grades and each teacher for each intelligence quotient level of students. When the teachers' norms were compared to the school norms, teachers' means were found to vary below and above the school means. The first null hypothesis that teachers instruct with consistent effectiveness throughout all student ability levels was, therefore, rejected.

A procedure to determine over-all effectiveness for each third of the ability range was applied and the results recorded. A questionnaire concerning the teacher's choice of ability level was administered and the results of this procedure tabled. These two variables were then fitted into a five-by-five matrix. A chi square test was performed based on this "observed" matrix and a calculated "expected" matrix. The resultant chi square was employed within a formula to produce a contingency coefficient "C." This coefficient was found significant at a confidence level greater than .05, thus rejecting the second null hypothesis that teacher desire for the instruction of a preferred ability level of students would not be an indicator of an area of teacher efficiency measured by the norm comparison procedure.

CHAPTER V

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

Summary of Procedures and Findings

Students who were in the second or the third grade in the Duncan, Oklahoma, Public School System during the school years 1966-67, 1967-68, 1968-69, and 1969-70 were used as the basis for measurement in this study. Student eligibility was determined by the possession of two scores--an intelligence quotient and an achievement test composite score--and by attendance in a single teacher's room for at least 100 days out of the school year. A total of 1063 students in the second grade and 1091 students in the third grade was found eligible.

On the basis of their intelligence quotients, the students in each grade were self-selected into ability levels of five-point ranges of intelligence quotient. All the grade equivalent scores of the achievement test within each ability level were summed, and a mean was computed. This procedure produced school norms in each grade for the range of ability levels.

The teachers in the system who had taught either the second grade or the third grade exclusively for the entire

four-year period were considered eligible for the research. A total of eleven teachers from a population of forty-three instructors was used as subjects.

The pupils in each of the eleven teachers' rooms over the four-year period were extracted from the total population of students. The ability level categorization and mean development procedure were accomplished for each teacher's pupils, providing norms for each teacher.

Each ability level mean of all the teachers was compared with the school norm. Notation was made if the teacher mean exceeded the school mean. On the basis of the apparent variation of effectiveness among the diverse ability levels, the first null hypothesis that teachers taught all ability levels with equal effectiveness was rejected.

For the latter part of the study, the students in each grade were divided into thirds based on the range of mental abilities. A procedure was devised to consolidate effectiveness in each of the ability levels within each third into a single measure of effectiveness for that third. If an instructor exceeded the school mean in 50 per cent or more of the ability levels in which he had students in a third, he was considered effective in that third. A table was drawn showing the effective thirds of all the instructors.

An individual interview was held with each eligible teacher. A check sheet guaranteed that certain points were made in the interview as a control. A single five-choice

questionnaire was verbally given to each teacher. This questionnaire allowed the teacher to express a desire to teach the upper, middle, or lower ability level of his grade. In addition, two other choices could have been made if he preferred a combination of all levels grouped together or none of the preceeding four choices for any reason. The results were tabled for all the teachers together.

A five-by-five matrix was formed by allowing the columns to represent the five possible choices on the questionnaire and the rows to depict the levels of effectiveness as indicated by the norm tables. A procedure was devised for entry into the "observed" matrix. From the column and row totals, an "expected" or theoretical matrix was drawn. With these two matrices, a chi square test was performed. The result of this statistic formed the basis for the final test--a contingency coefficient "C." This correlation was found to be significant in excess of the .05 confidnece level, thus rejecting the second null hypothesis that teacher's choice of instructional ability level does not appear to be a predictor of effectiveness with that level as indicated by the measurement of his pupils' academic achievement.

Conclusions

The rejection of the first null hypothesis appears to indicate that, in this school system, the concept of individual differences is found among persons other than students in

the educational situation. Capacities to teach a variety of ability levels seems to differ within individuals.

The significance of the correlation allowed rejection of the second null hypothesis. Based on this correlation, the discussion with the instructor in this investigation could effectively replace the norm comparison procedure as an indicator of the teacher's apparent level of effective instruction.

Implications

There exists evidence that the teachers involved in this research cannot be assumed equally effective with all ability levels of students. Methods of assignment for these teachers should take into consideration the individual areas of apparent effectiveness. Any other assignment procedure might be less beneficial to those students involved.

The norm comparison procedure could serve as a tool to match most efficiently the student's ability level with the teacher's effectiveness level. For the teachers involved in this study, this idea could lend support in the search for aids to individualized instruction. Teachers could be described as "good" with those categories of their students which have shown superior development. If all individuals are equally worthy of quality education, no stigma could be attached to the effective instruction of any ability level grouping.

The correlation of the interview response with the norm comparisons would appear to indicate that consultation concerning assignment might be wise for those involved in this

project. It would seem that the effectiveness of grouping children could be enhanced if the desires of this faculty are honored.

For the employment of the norm comparison procedure, several factors should be considered to produce maximum results from its use. This process demands at least two teachers per grade level in the elementary school or two teachers per grade level and subject area in the departmentalized secondary school.

The procedure could be used to test a teacher's effectiveness with a class of homogeneously grouped students if he has had that level for the total period of time. However, to indicate best a teacher's effective levels from among all the levels, grouping should be heterogeneous and random. Homogeneous grouping, wherein a system of rotation had allowed the teacher to instruct all ability levels of students within the testing period, would be equally useful.

Departmentalization at any level makes the procedure slightly more complicated. Composite scores on the test of achievement could not be used. Of course, more than one teacher must teach the subject for the employment of the test.

To ease calculation of the data, achievement tests should be given at approximately the same time each year. Grade equivalent scores would then be comparable. Otherwise interpolation to equalize the dates of testing would be necessary.

Within this study exists no attempt to perform a judgment of over-all comparison of the teachers involved. No effort to rate or rank teachers was included. A statement that a fine teacher of upper-mental-ability children was "a better teacher" than the outstanding instructor of low-mental-ability children would be impossible to make from these proceedings. Both teachers should be considered excellent.

With this procedure, no comparisons across school boundaries can be accomplished. Only within the confines defined by the data analyzed may effectiveness be considered. The situational nature of the results of such a project must be emphasized.

A certain amount of analysis of the responses of teachers should be performed. Referring to Table 3, page 48, it can be seen that several teachers were shown effective in more than one third. In fact, only two teachers were shown outstanding in a single level. Table 3 shows that those who chose only a single area were still able to choose their effective level.

It appears that those having a desire for no single level of ability tend to be shown effective in no area. Two persons chose the combination of all levels and were found ineffective in all levels. Within this study, employment of teachers who express a single ability level could possibly produce the most effective staff.

Of those people in the study, there were none who indicated they preferred a grade other than those being investigated. All teachers stated that they preferred their grade to all other grades.

A closely related observation was that no teacher checked the fifth choice--"none of the levels mentioned." However, the norm comparison procedure indicated that three of the teachers should have chosen this response. If the level a teacher chose did not result in a match, he was found to have no effective third.

All teachers appeared quite interested and enthusiastic concerning the concepts involved in this study. Every instructor expressed agreement that teachers are as individual as children.

With the exception of the two teachers who chose answer four--"a combination room of all three ability levels"--the teachers expressed definite desires to instruct homogeneously grouped classes. Several teachers made the comment that when they were teaching grouped classes, in the past, they wished for random assignment of students. However, after heterogeneous grouping was restored, they found that they definitely liked the homogeneous grouping better.

Need for Further Research

Replicate studies would be worthwhile in that they could be concerned with the experimental as well as the

numerical expansion of the elements of this exploratory research. A numerical increase could be built up by a series of analyses. Various factors, such as the period of research, could be experimentally examined to find the point of optimum returns.

Later research might investigate other variables among school systems which might affect the results. The procedure could be executed in various school systems of different sizes, socio-economic condition, rural or urban nature, and other such variables. A theory base would need to be developed to give justification for hypothesizing that these variables could affect the research results.

Another aspect of this work which might be analyzed further would be the tests of teacher effectiveness using a variety of subject area subtest scores rather than the composite achievement scores. From the point of view of departmentalization in the elementary schools, this procedure might provide some justification for placement by subject area as well as ability level. Subject areas might form the basis for the use of this procedure on the secondary level.

Comparative studies involving the results of tests of achievement between systems employing this form of placement procedure and systems using random placement should be devised. This examination could provide evidence bearing on the actual outcome of such procedures.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ackerman, William I. "Teacher Competence and Pupil Change." The Harvard Educational Review, XXIV (Fall, 1954), 273-89.

Anastosi, Anne and Foley, John P., Jr. Differential Psychology. New York: The Macmillan Company, 1949.

Angoff, William H. "Measurement and Scaling." The Encyclopedia of Educational Research. 3rd ed. Edited by Chester W. Harris. New York: The Macmillan Company, 1960.

Barr, A. S. "The Measurement and Prediction of Teacher Efficiency: A Summary of Investigations." Journal of Experimental Education, XVI (June, 1;48), 203-83.

_____; Worcester, D. A.; Abell, Allan; Beecher, Clarence; Jensen, Leland E.; Peronto, Archie L.; Ringness, Thomas A.; and Schmid, John. Wisconsin Studies of the Measurement and Prediction of Teacher Effectiveness: A Summary of Investigations. Madison, Wisconsin: Dembar Publications, Inc., 1961.

Biddle, Bruce J. and Ellena, William J., eds. Contemporary Research on Teacher Effectiveness. New York: Holt, Rinehart and Winston, 1964.

Bolton, Floyd B. "Evaluating Teaching Effectiveness Through the Use of Scores on Achievement Tests." Journal of Educational Research, XXXVIII (May, 1945), 691-96.

Brink, William G. "Secondary Education--Programs." The Encyclopedia of Educational Research. 3rd ed. Edited by Chester W. Harris. New York: The Macmillan Company, 1960.

Brown, Ralph R. "The Time Interval between Test and Re-test in its Relation to the Consistency of the Intelligence Quotient." The Journal of Educational Psychology, XXIV (February, 1933), 81-96.

Burks, Barbara Stoddard. "The Relative Influence of Nature and Nurture upon Mental Development: A Comparative Study of Foster Parent-Foster Child Resemblance and True Parent-True Child Resemblance." Nature and Nurture. Twenty-Seventh Yearbook of the National Society for the Study of Education, Part I. Bloomington, Ill.: Public School Publishing Company, 1928.

_____. "A Summary of Literature on the Determiners of the Intelligence Quotient and the Educational Quotient." Nature and Nurture. Twenty-Seventh Yearbook of the National Society for the Study of Education, Part II. Bloomington, Ill.: Public School Publishing Company, 1928.

Burton, William H. and Brueckner, Leo J. Supervision. New York: Appleton-Century-Crofts, 1955.

Combs, Arthur W. The Professional Education of Teachers. Boston: Allyn and Bacon, Inc., 1965.

Committee on the Criteria of Teacher Effectiveness. Report of the Committee. Review of Educational Research, XXII (June, 1952), 238-63.

Day, L. C. "The Teaching Quotient." The Elementary School Journal, XXXIII (April, 1933), 604-07.

Davis, Donald E. and Nickerson, Neal C. Critical Issues in School Personnel Administration. Chicago: Rand McNally and Company, 1968.

Domas, Simeon J. and Tiedeman, David U. "Teacher Competence: An Annotated Bibliography." Journal of Experimental Education, XIX (December, 1950), 101-219.

Evans, Kathlee M. "A Critical Survey of Methods of Assessing Teaching Ability." British Journal of Educational Psychology, XXI (June, 1951), 89-95.

Hall, Roy M. "Staff--Selection and Appointment." The Encyclopedia of Educational Research. 3rd ed. Edited by Chester W. Harris. New York: The Macmillan Company, 1960.

Heathers, Glen. "Grouping." The Encyclopedia of Educational Research. 4th ed. Edited by Robert L. Ebel. London: The Macmillan Company, 1969.

Jenkins, Albian U. "Measurement of Teaching Efficiency by Means of Standardized Tests." Department of Elementary Schools Principals Bulletin, VIII (1929), 373-82.

Jones, E. S. "Suggestion for Teacher Measurement." School and Society, VI (1971), 321.

Lancelot, William H.; Barr, Arvil S.; Torgerson, Theodore L.; Johnson, Carl E.; Lyon, Vergil E.; Wolvoord, Anthony C.; and Betts, Gilbert Lee. The Measurement of Teaching Efficiency. New York: The Macmillan Co., 1935.

McCall, William A. How To Measure in Education. New York: The Macmillan Company, 1922.

_____. Measurement of Teacher Merit. Raleigh, N.C.: Department of Public Instruction, 1952.

Matthew, Janet and Luckey, Bertha. "Notes on Factors that May Alter the Intelligence Quotient in Successive Examinations." Nature and Nurture. Twenty-Seventh Yearbook of the National Society for the Study of Education, Part I. Bloomington, Ill.: Public School Publishing Company, 1928.

Michael, William B. "Aptitudes." The Encyclopedia of Educational Research. 3rd ed. Edited by Chester W. Harris. New York: The Macmillan Company, 1960.

Mitzel, Harold E. "Teacher Effectiveness." The Encyclopedia of Educational Research. 3rd ed. Edited by Chester W. Harris. New York: The Macmillan Company, 1960.

Nelson, Martin J. "Intelligence and Special Aptitude Tests." The Encyclopedia of Educational Research. 4th ed. Edited by Robert L. Ebel. London: The Macmillan Company, 1969.

Otis, Arthur S. and Lennon, Roger T. Otis-Lennon Mental Ability Test, Form J. New York: Harcourt, Brace and World, Inc., 1967.

Ragan, William B. Modern Elementary Curriculum. 3rd ed. New York: Holt, Rinehart and Winston, 1966.

Rogers, Agnes L., Durling, Dorothy, and McBride, Katharine. "The Effect on the Intelligence Quotient of Change from a Poor to a Good Environment." Nature and Nurture. Twenty-Seventh Yearbook of the National Society for the Study of Education, Part I. Bloomington, Ill.: Public School Publishing Company, 1928.

Seyfert, Warren C. and Tyndal, Balfour S. "An Evaluation of Differences in Teaching Ability." Journal of Educational Research, XXVIII (September, 1934), 10-15.

Stephens, J. M. and Lichtenstein, Arthur. "Factors Associated with Success in Teaching Grade Five Arithmetic." Journal of Educational Research, XL (May, 1947), 683-94.

Terman, Lewis M. The Intelligence of School Children. Boston: Houghton Mifflin Company, 1919.

_____. The Measurement of Intelligence. Boston: Houghton Mifflin Company, 1916.

Thelen, Herbert A. "Grouping for Teachability." Theory into Practice. II (1963), 81-89.

Tiegs, Ernest W. and Clark, Willis W., devisors. California Achievement Tests Complete Battery and Separate Reading, Arithmetic, and Language Tests. 1957 ed. Monterey, Calif.: California Test Bureau, 1957.

Tomlinson, Loren R. "Recent Studies in the Evaluation of Teaching." Educational Research Bulletin, XXXIV (October, 1955), 172-86 and 196.

Trump, J. Lloyd and Miller, Delmas F. Secondary School Curriculum Improvement. Boston: Allyn and Bacon, Inc., 1968.

Tuddenham, Read D. "Intelligence." The Encyclopedia of Educational Research. 4th ed. Edited by Robert L. Ebel. London: The Macmillan Company, 1969.

Van Zwoll, James A. School Personnel Administration. New York: Appleton-Century-Crofts, 1964.

Watters, William A. "Annotated Bibliography of Publications Related to Teacher Evaluation." Journal of Experimental Education, XXII (June, 1954), 351-69.

Woodworth, R. S. Heredity and Environment. New York: Social Science Research Council, 1941.

APPENDICES

I. CHECKLIST FOR THE INTERVIEW

II. SCHOOL NORMS

III. TEACHER'S MEANS

APPENDIX I

CHECKLIST FOR THE INTERVIEW

Several points were mentioned in all of the interviews before the administration of the questionnaire. The following considerations were used as a checklist for control.

All instructors were informed concerning the philosophy of the work by describing the condition of the realization of individual differences in students as evidenced by grouping practices. However, little consideration is often given to the idea that teachers are also individuals. Their feelings or reactions were then solicited.

Assurances of anonymity and a statement that no one else would be consulted concerning his reaction was given to the teacher. Described as "ivory tower" or purely theoretical research, this study was portrayed as having no effect on the teacher.

A brief statement was made concerning the apparent social stigma which has been attached to instruction of the low ability students. Mention was made of the current interest in "special education" and of the ability of these students to learn. Additionally, it was suggested that some teachers seem to think that the instruction of the high ability groups demands less effort. However, there was an oral reminder that problems may exist at all levels.

Following the questionnaire, the teacher was given the original of all notes taken and his response to the question. A carbon copy was retained by the investigator. As a control device, the teachers were asked to refrain from discussing the interview for that period of time deemed necessary to complete all interviews.

APPENDIX II

SCHOOL NORMS

TABLE 5a

GRADE 2

Intelligence Quotient Level	Number of Students	Mean Score of the Level	Standard Deviation
Above 134	13	4.438	.545
130-134	26	4.180	.394
125-129	40	4.198	.205
120-124	71	4.178	.298
115-119	94	4.008	.348
→110-114	119	3.873	.386
105-109	149	3.784	.292
100-104	164	3.639	.237
→ 95-99	132	3.473	.315
90-94	94	3.398	.335
85-89	75	3.305	.237
80-84	45	3.042	.436
75-79	26	3.077	.335
70-74	10	2.740	.561
Below 70	6	2.983	.668
Total	1063		

TABLE 5b

GRADE 3

Intelligence Quotient Level	Number of Students	Mean Score of the Level	Standard Deviation
Above 134	18	4.861	.110
130-134	37	4.816	.290
125-129	56	4.739	.134
120-124	87	4.691	.226
115-119	111	4.548	.239
→ 110-114	122	4.509	.263
105-109	148	4.446	.158
100-104	148	4.364	.300
→ 95-99	130	4.226	.286
90-94	90	4.141	.078
85-89	69	3.923	.398
80-84	40	3.670	.466
75-79	19	3.821	.399
70-74	13	3.338	.612
Below 70	3	3.533	.927
Total	1091		

APPENDIX III

TEACHER'S MEANS

TABLE 6a

TEACHER A--GRADE 2

Intelligence Quotient Level	Number of Students	Mean Score of the Level	Exceed School Mean	Exceed ± One Standard Deviation
Above 134	1	4.500	x	
130-134	1	4.400	x	
125-129	5	4.360	x	
120-124	5	3.900		
115-119	4	4.175	x	
→110-114	5	3.860		
105-109	5	3.620		
100-104	14	3.600		
→ 95-99	7	3.485	x	
90-94	4	3.675	x	
85-89	2	3.300		
80-84	4	3.125	x	
75-79	0	. .		
70-74	2	2.300		
Below 70	0	. .		
Total	59			

TABLE 6b

TEACHER B--GRADE 2

Intelligence Quotient Level	Number of Students	Mean Score of the Level	Exceed School Mean	Exceed ± One Standard Deviation
Above 134	0	. .		
130-134	2	3.750		*
125-129	1	3.900		*
120-124	1	4.400	x	
115-119	3	3.933		
→110-114	8	3.825		
105-109	9	3.766		
100-104	14	3.550		
→ 95-99	20	3.390		
90-94	7	3.200		
85-89	2	3.600	x	
80-84	0	. .		
75-79	3	3.067		
70-74	0	. .		
Below 70	2	2.700		
Total	72			

TABLE 6c

TEACHER C--GRADE 2

Intelligence Quotient Level	Number of Students	Mean Score of the Level	Exceed School Mean	Exceed ± One Standard Deviation
Above 134	1	4.700	x	
130-134	1	4.500	x	
125-129	1	4.700	x	*
120-124	3	4.233	x	
115-119	5	4.120	x	
→110-114	10	3.780		
105-109	11	3.882	x	
100-104	11	3.836	x	
→ 95-99	5	3.680	x	
90-94	4	3.325		
85-89	3	3.500	x	
80-84	1	3.100	x	
75-79	0	. .		
70-74	1	2.800	x	
Below 70	0	. .		
Total	57			

TABLE 6d

TEACHER D--GRADE 2

Intelligence Quotient Level	Number of Students	Mean Score of the Level	Exceed School Mean	Exceed ± One Standard Deviation
Above 134	0	. .		
130-134	3	4.333	x	
125-129	3	4.500	x	
120-124	4	4.175		
115-119	7	4.271	x	
→110-114	12	3.942	x	
105-109	10	3.860	x	
100-104	11	3.755	x	
→ 95-99	12	3.683	x	
90-94	8	3.575	x	
85-89	5	3.180		
80-84	8	3.137	x	
75-79	6	3.383	x	
70-74	0	. .		
Below 70	1	2.400		
Total	90			

TABLE 6e

TEACHER E--GRADE 2

Intelligence Quotient Level	Number of Students	Mean Score of the Level	Exceed School Mean	Exceed ± One Standard Deviation
Above 134	1	4.100		
130-134	0	. .		
125-129	1	3.900		
120-124	5	4.180	x	
115-119	1	4.100	x	
→110-114	6	3.500		
105-109	11	3.664		
100-104	11	3.391		*
→ 95-99	13	3.462		
90-94	4	3.350		
85-89	13	3.269		
80-84	9	3.044	x	
75-79	3	3.000		
70-74	1	1.800		*
Below 70	1	3.300	x	
Total	80			

TABLE 6f

TEACHER F--GRADE 3

Intelligence Quotient Level	Number of Students	Mean Score of the Level	Exceed School Mean	Exceed ± One Standard Deviation
Above 134	2	4.850		
130-134	1	4.900	x	
125-129	13	4.777	x	
120-124	7	4.700	x	
115-119	10	4.500		
→110-114	9	4.400		
105-109	10	4.530	x	
100-104	12	4.300		
→ 95-99	13	4.054		
90-94	7	4.271	x	
85-89	5	3.880		
80-84	3	3.800	x	
75-79	1	3.800		
70-74	1	2.100		*
Below 70	0	. .		
Total	94			

TABLE 6g

TEACHER G--GRADE 3

Intelligence Quotient Level	Number of Students	Mean Score of the Level	Exceed School Mean	Exceed ± One Standard Deviation
Above 134	2	4.850		
130-134	4	4.950	x	
125-129	4	4.950	x	*
120-124	7	4.914	x	
115-119	4	4.725	x	
→110-114	8	4.562	x	
105-109	15	4.680	x	*
100-104	10	4.310		
→ 95-99	17	4.259	x	
90-94	11	4.091		
85-89	8	4.062	x	
80-84	4	3.700	x	
75-79	2	3.900	x	
70-74	1	3.300		
Below 70	0	. .		
Total	97			

TABLE 6h

TEACHER H--GRADE 3

Intelligence Quotient Level	Number of Students	Mean Score of the Level	Exceed School Mean	Exceed ± One Standard Deviation
Above 134	0	. .		
130-134	3	4.833	x	
125-129	1	4.900	x	*
120-124	3	4.833	x	
115-119	6	4.750	x	
→110-114	13	4.685	x	
105-109	9	4.655	x	*
100-104	21	4.571	x	
→ 95-99	14	4.436	x	
90-94	10	4.390	x	*
85-89	2	4.350	x	*
80-84	4	4.225	x	*
75-79	1	4.400	x	*
70-74	2	4.250	x	*
Below 70	1	4.500	x	*
Total	90			

TABLE 6i

TEACHER I--GRADE 3

Intelligence Quotient Level	Number of Students	Mean Score of the Level	Exceed School Mean	Exceed ± One Standard Deviation
Above 134	1	4.900	x	
130-134	7	4.800		
125-129	2	4.750	x	
120-124	13	4.662		
115-119	16	4.638	x	
→110-114	10	4.480		
105-109	9	4.466	x	
100-104	6	4.450	x	
→ 95-99	8	4.350	x	
90-94	6	4.066		
85-89	3	4.033	x	
80-84	0	. .		
75-79	0	. .		
70-74	0	. .		
Below 70	0	. .		
Total	81			

TABLE 6j

TEACHER J--GRADE 3

Intelligence Quotient Level	Number of Students	Mean Score of the Level	Exceed School Mean	Exceed ± One Standard Deviation
Above 134	1	5.000	x	*
130-134	2	4.650		
125-129	2	4.850	x	
120-124	8	4.750	x	
115-119	8	4.562	x	
→ 110-114	12	4.483		
105-109	13	4.408		
100-104	11	4.309		
→ 95-99	16	4.263	x	
90-94	6	4.233	x	*
85-89	8	3.725		
80-84	8	3.750	x	
75-79	4	3.800		
70-74	2	3.000		
Below 70	0	. .		
Total	101			

TABLE 6k

TEACHER K--GRADE 3

Intelligence Quotient Level	Number of Students	Mean Score of the Level	Exceed School Mean	Exceed ± One Standard Deviation
Above 134	0	. .		
130-134	2	4.900	x	
125-129	2	4.700		
120-124	4	4.500		
115-119	7	4.528		
→110-114	8	4.300		
105-109	11	4.300		
100-104	6	4.433	x	
→ 95-99	9	4.222		
90-94	3	4.100		
85-89	7	3.971	x	
80-84	3	3.400		
75-79	2	3.850	x	
70-74	1	3.300		
Below 70	0	. .		
Total	65			