THE TESTS OF BASIC EXPERIENCES AND

THE SCREENING TEST FOR ASSIGNMENT OF

REMEDIAL TREATMENTS AS PREDICTORS OF

READINESS FOR FIRST GRADE

by

Darryl G. Gibson

Approved by:

Richard Levin

Dr. Richard Levin, Chairman, Thesis Committee

W. T. Snipes

Dr. Thomas Snipes, Member, Thesis Committee

Frank Terrant

Dr. Frank Terrant, Member, Thesis Committee

Joyce Crouch

Dr. Joyce Crouch, Chairman, Department of Psychology

Richard H. Rupp

Dr. Richard Rupp, Dean of the Graduate School

THE TESTS OF BASIC EXPERIENCES AND

THE SCREENING TEST FOR ASSIGNMENT OF

REMEDIAL TREATMENTS AS PREDICTORS OF

READINESS FOR FIRST GRADE

———————

A THESIS

PRESENTED TO

THE FACULTY OF THE GRADUATE SCHOOL

APPALACHIAN STATE UNIVERSITY

———————

IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE

MASTER OF ARTS IN

CLINICAL PSYCHOLOGY

———————

by

Darryl G. Gibson

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

The present study was designed to test the ability of scores on the Tests of Basic Experiences and the Screening Test for Assignment of Remedial Treatments to identify pupils at the beginning of the kindergarten year who would be identified as not ready for first grade by the Metropolitan Readiness Tests score at the end of the kindergarten year. Two hundred students were administered three tests and scores were inter correlated using the Pearson Product Moment formula. Correlations were high among most relationships. Results are discussed in terms of their implications for early remediation as special assignments based on results of Fall tests.

INTRODUCTION

Over the past decade, nearly every aspect of education has been exposed to some type of research and/or rennovation. One of the major actions to result from all of the attention directed toward education has been an attempt to meet the needs of all students, which takes on special meaning when applied to students with educational handicaps and/ or disabilities. Meeting the needs of these students often requires special programs, classes and instruction. Federal, state and in many cases, local governments have adopted legislation and appropriated funds to assist in establishing the necessary facilities and programs.

Initially, placement in these programs came only after a student had exhibited several years of poor academic performance and below grade level achievement in a regular classroom setting. However, educators, psychologists and medical professionals began to request and support the early identification of students with learning problems (Keogh & Becker, 1973), arguing that the effectiveness of programs of remediation and special instruction were enhanced by the identification and treatment of these problems as early as possible in the educational and developmental process (Rogolsky, 1969). Subsequently, attention was directed toward preschool and kindergarten screening of all students, thereby enabling early identification of high risk students [a prediction that a student will fail or have problems in the existing school program.(Keogh & Becker, 1973)].

Opinions vary as to what provides the best early indication of later school achievement; subsequently, research and screening and readiness

tests based on teacher predictions (Kirk, 1966 and Wang, 1974), behavior
rating scales (Correll, 1962, and Attwell, Oppert and Meyers, 1967 and
Spivack and Swift, 1968), perceptual-motor skills (Frostig, Maslow,
Lefever and Whittlesey, 1963, Keogh and Smith, 1960 and Koppitz, 1964),
cognitive functioning (Ahr, 1967 and Harris, 1963), language (deHirsch,
Jansky and Langford, 1966, Durkin, 1960 and Faust, 1970) and preschool
experience (Bronfenbrenner, 1965 and Hubrrty and Swan, 1974) have all
had an effective part in early identification. However, since many of
these instruments have questionable validity, frequently fail to provide
the classroom teacher with data for direct remedial programming and are
often economically impractical for screening large numbers of children,
they have experienced limited application on an individual basis but have
been combined and included as subtests in screening instruments designed
to meet the previously mentioned criteria. Such a practice also supports
Rogolsky's (1969) observation that since the etiology of learning problems
remains vague, screening programs should use a variety of measures which
tap visual perceptive, spatial and verbal fields.

Although educators strongly support screening and early identifica-
tion programs, they also question the validity of tests used as predictors,
caution against the possibility of a "self-fulfilling prophecy" in that
what is actually being done is hypothesizing and not identifying, and
finally they critically review the routine collection of data and failure
to modify programs identified as failure producing for a child prior to
placing the child in the program (Keogh and Becker, 1973 and Rogelsky,
1969).

This study attempted to address some of the final concerns expressed by Keogh and Becker (1973) and Rogolsky (1969). It was designed to determine how well two group screening instruments administered at the beginning of the kindergarten year predicted readiness for first grade as measured by a third instrument administered at the end of the kindergarten year. Several questions were investigated by the study:

1. Can the Screening Test For Assignment to Remedial Treatments (START) (Ahr, 1968) identify kindergarten pupils at the beginning of the kingergarten year who will not be ready for first grade according to Metropolitan Readiness Tests (MRT) (Hildreth, Griffiths and McGauvran, 1969) scores at the end of the kindergarten year?

2. Can the Tests of Basic Experiences (TOBE) Moss, 1970-71) identify kindergarten pupils at the beginning of the kindergarten year who will not be ready for first grade according to MRT scores at the end of the kindergarten year?

3. In as much as Faust (1970) emphasized that individual characteristics change as a function of interaction with the environment and that there are, thus, few "inherent, stable traits of the individual" which allow long term prediction, will there be a statistically significant difference between the scores of pupils identified by the START as needing remedial instruction at the beginning of the kindergarten year and these same students at the end of the kindergarten year, although they received no remediation?

4. What is the correlation between the TOBE's prediction of readiness

and the START's prediction of readiness at the beginning of the
kindergarten year?

METHOD

Subjects

Two hundred pupils, ranging in age from five years, zero months to
six years, six months from ten kindergarten classrooms in the Winston-
Salem/Forsyth County School System in North Carolina served as subjects
in this study.

The Winston-Salem/Forsyth County School System was composed of 69
schools, grades K-12, with a pupil population of 50,000. Approximately
1,250 of these students were enrolled in 50 kindergarten classrooms,
having an average enrollment of 25 pupils per classroom.

The population was nearly evenly divided between male and female
with the percentages being 46.5% and 53.5% respectively. White pupils
composed 65% of the study population while 35% was black. This latter
percentage was very compatible to the school system's population and the
general population in the Winston-Salem/Forsyth County area(Tirrell, 1976).

Apparatus

Screening Test for the Assignment of Remedial Treatments

Edward A. Ahr was active in research and screening among kinder-
garten and preschool students long before he developed the START. Prior
to the development of the START, he developed the Screening Test of
Academic Readiness (STAR) (Ahr, 1966) for use with kindergarten and pre-
school students. Recognizing the need for a more definitive method of

determining areas of weakness and assignment to remedial groups, Ahr (1967) began compiling the various skills his and other research indicated as valid in assessing readiness. The START is the cumulative result of these efforts.

Description: The START is a group test for children between the ages of four years, six months and six years, five months. The test consists of four subtests (Visual Memory, Auditory Memory, Visual Copying and Visual Discrimination) and may be administered in one hour and scored in six minutes by an untrained examiner. The START is recommended for teacher use as a measure of a child's developmental level in each skill area to assist in planning individualized instruction.

Standardization: The START was standardized on 500 middle class, nursery school and kindergarten children between the ages of four years, six months and six years, five months. These children were drawn from urban, suburban and semi-rural areas. No further description of the subjects was given.

Validity: Estimates of validity, using a teacher rating scale as a criterion, ranged between .45 and .64 for individual subtest scores with a coefficient of .70 reported for total test scores.

No studies were reported concerning the validity of the START using achievement measures as criteria and no studies examining the usefulness of START results for planning remedial instruction were reported.

Reliability: Reliability of the START was reported as uniformly high, ranging from .92 to .97 for total test scores and between .78 and .93 for the various subtest scores. The standard error of measurement,

depending on the reliability measure used, ranged between one and two raw score points for total START scores.

Metropolitan Readiness Tests

The Metropolitan Readiness Tests were first developed in 1933 and have undergone several revisions since that time. The authors recognized linguistic attainments and aptitudes, visual and auditory perception, muscular coordination and motor skills, number knowledge, and the ability to follow directions and to pay attention in group work as among the chief factors contributing to readiness for beginning schoolwork (Hildreth, Griffiths and McGavron, 1969). In conjunction with this was the recognition that the degree of advancement in these skills also depended upon other physical, intellectual, social, emotional and other environmental factors. Metropolitan Readiness Tests were designed to measure the extent to which these skills and abilities were developed in students preparing to enter first grade (Hildreth, Griffiths and McGavron, 1969).

Description: The Metropolitan Readiness Tests (MRT's) are designed for testing pupils at the end of the kindergarten year or beginning of first grade. The tests are constructed to measure some of the chief factors contributing to readiness for beginning schoolwork such as linguistic attainments and aptitudes, visual and auditory perception, muscular coordination and motor skills, number knowledge, and the ability to follow directions. The MRT's may be administered by the classroom teacher to groups of pupils and all tests (subtests) may be given in one hour.

The six tests included in the MRT's are the following:

Test 1. Word Meaning

Test 2. Listening

Test 3. Matching

Test 4. Alphabet

Test 5. Numbers

Test 6. Copying

(A seventh test, Draw-a-Man, is optional.)

Standardization: The current revision of the MRT's, Form A, was standardized in 1964 on a sample of 15,000 pupils in 70 school systems. Form B norms were established by means of equating results on this form with those on Form A for the same subjects. A comparison of results for the two forms showed they were equivalent in terms of total score. For three of the subtests, there existed a one-point difference along a portion of the score scale.

Pupils selected in the standardization sample were drawn from twelve states (including North Carolina) and five geographical regions. The subjects were from urban, suburban and rural areas and from various minority groups and socioeconomic levels. An equal number of boys and girls, with an age range from 5-7 years to 7-10 years and IQ range of 60 to 150, were tested.

Validity: The correlation between the MRT's scores for kindergarten pupils and Stanford Achievement Test: Primary I test scores for 9,497 subjects was at a level of .65 for total overall achievement on the six Stanford and Metropolitan subtests (Stanford was administered in October of Grade 1).

Separate studies carried out with 421 first grade pupils (88 in all Negro schools) and 119 first grade pupils, using the Metropolitan Achievement Tests as the predicted criterion, yielded correlations of .67 and .69, respectively, between total readiness and achievement scores. A number of other studies of the MRT's total scores and Metropolitan and Stanford Achievement subtest scores showed correlations ranging from .47 to .67. These studies employed sample sizes of 74 to 553.

Reliability: Split-half reliabilities ranged from .91 to .94 for total test scores in three separate studies of first grade pupils. Four additional studies of kindergarten pupils with sample sizes of 59 to 88 yielded split-half reliabilities of .90 to .95 for total test scores with the measurement error of an individual total score ranging from 3 to 5 points. For these subjects the Alphabet test showed the highest degree of reliability (.88). The Matching test followed with a median reliability value of .82, then Copying with .81, Numbers with .80, and Word Meaning and Listening were the least reliable with median values of .61 and .52, respectively.

Tests of Basic Experiences General Concepts   Level K

As is frequently the case with many research instruments and tests, the TOBE are the result of several revisions of previous instruments of the same general theme. The TOBE's predecessor, the Test of Basic Information was developed in 1961 as part of a study designed to determine whether mentally retarded children in special classes made more progress than similar children left in regular classes (Moss, 1968). The test was composed of two parts, Part I which measured non-academic know-

knowledge and Part II which contained many reading factors and correlated highly with ability to read. The first administration of the TOBI revealed that Black retarded children scored substantially below children in all other groups. Following the revision a year later, these differences were not noted (Moss, 1968).

Margaret Moss (1968) began development of the TOBE using items from the TOBI and Items of Space and Location (ISL) which was used in research for her dissertation. Unlike the two previous instruments (TOBI and ISL), the TOBE places primary emphasis on experiences (concept formation) rather than upon information (facts) (Moss, 1968).

Description: The TOBE - Level K is a series of five tests designed for children in the preschool or kindergarten age group. The five tests are Mathematics, Language, Science, Social Studies and General Concepts which consists of several items from the other four areas (Mathematics-6, Language-5, Science-6, and Social Studies-4). The General Concepts Test is a more general measure assessing the richness of a student's background, experiences and familiarity with various concepts. (The General Concepts Test will be used in this study.) It requires approximately 25 minutes to administer and makes use of proctors to ensure that the requirements of a valid testing session are met. The TOBE can be hand-scored with relative ease using answer keys.

Standardization: The TOBE were standardized on approximately 10,300 children from all types of public or private prekindergartens, kindergartens, first grades or second grades. The population was not a probability sample but consisted of 422 classes in 145 schools in 44 cities. Subjects were drawn from the East, South, Midcontinent and West

and from inner-city, urban, suburban and small city communities.

Validity: There were no numerical estimates of validity with the TOBE for reported lack of an appropriate criterion measure. However, a content validation study employing TOBE classification of subject items as compared to teacher classification yielded high percentage of agreement.

Reliability: Reliability for the TOBE was relatively high ranging from .84 to .79. The standard error of measurement ranged from 2.29 to 2.39 raw score points on individual tests.

Procedure

A meeting was held with all kindergarten teachers and principals at the beginning of the school year to explain the study and request their cooperation. Twelve classrooms were randomly selected from among co-operating schools to participate in the study.

During the first week in October, pupils in each of the classrooms were administered the START by a volunteer certified teacher, who was trained in the administration of this instrument. Test protocols were collected and scored.

During the same time period (first-second week in October) individual classroom teachers administered, scored and recorded the results of the TOBE.

Finally all classrooms were again administered the START and the Metropolitan Readiness Tests the following May by volunteer certified teachers and school psychologists.

RESULTS

Scores obtained on the four administrations were recorded and computed using Raw Score totals. Although each test provided for the Raw Score totals to be converted to percentile ranks and stanines, to facilitate statistical computations and maintain as much accuracy and sensitivity as possible, all scores were retained as raw scores.

The six variables in the study (race, sex, TOBE, START 1, START 2 and MRT) were subjected to several statistical computations to evaluate their inter-relationships. Table 1 contains the total score ranges, means and standard deviations for each of the tests.

Correlation coefficients between the six variables and their statistical significances are shown in Table 2. The higher correlations (MRT/START = .7428, MRT/TOBE = .7352, START 1/START 2 = .6544, START 1/MRT - .6322, TOBE/START 2 = .6120), all of which were significant at the .001 level of confidence, provided direct responses to the four questions posed in the Introduction section of this paper.

The START 1/MRT coefficient (.6322) and the TOBE/MRT coefficient (.7352) provided affirmative answers to questions number 1 and 2 which asked if the START and the TOBE could identify pupils at the beginning of the kindergarten year who would not be ready for first grade according to MRT scores obtained at the end of the kindergarten year.

In response to question number 3 "...will there be a statistically significant difference between the scores of pupils identified by the

TABLE 1

STATISTICAL DATA ON THE FOUR TESTS IN STUDY

FOR THE COMPLETE SAMPLE (N = 200)

Test	Total Score Range	Mean	Standard Deviation
TOBE	3-26	15.99	5.48
START 1	2-71	40.74	17.66
START 2	29-73	60.62	7.74
MRT	30-96	66.92	12.29

TABLE 2

PEARSON CORRELATION COEFFICIENTS

	RACE	SEX	START 1	START 2	TOBE	MRT
RACE	-	-	-	-	-	-
	-	-	-	-	-	-
SEX	-0.0305	-	-	-	-	-
	S-0.668	-	-	-	-	-
START 1	-0.2846	0.0497	-	-	-	-
	S=0.001	S-0.485	-	-	-	-
START 2	-0.3424	0.1496	0.6544	-	-	-
	S=0.001	S=0.034	S=0.001	-	-	-
TOBE	-0.4204	-0.0182	0.5986	0.6120	-	-
	S=0.001	S=0.798	S=0.001	S=0.001	-	-
MRT	-0.3583	0.0769	0.6322	0.7428	0.7352	-
	S=0.001	S=0.279	S=0.001	S=0.001	S=0.001	-

S = probability level that coefficient is different from zero (0.000)

START as needing remedial instruction at the beginning of the kindergarten year and these same students at the end of the kindergarten year, although they received no instruction?", the correlation coefficient of .6544 between START 1/START 2 indicated that differences observed between the scores for these groups could be considered significant.

Finally, the correlation between the TOBE and START 1 produced a coefficient of .5986 which was also significant at the .001 level of confidence.

The correlation of race and sex with the four other variables in the study produced consistently low coefficients. However, to evaluate for correlations among individual racial - sexual groups (i.e., black male, white male, etc.) Pearson Product Moment Correlations were also computed for these groups and the resulting correlation coefficients are shown in Table 3.

DISCUSSION

These coefficients are generally consistent with the coefficients obtained from correlating the same variables in the total study population (i.e., TOBE/MRT = .7352, START 1/MRT = .6322). The correlation between the TOBE and the MRT for males was considerably higher (black males = .7603, white males = .7558, black females = .6348, white females = .6662) than the same correlations for females, however the correlation was consistent within individual sexual groups. Between the START 1 and the MRT, black females produced the highest correlation (.7063) with white females producing a very poor correlation (.4867).

TABLE 3

PEARSON PRODUCT - MOMENT CORRELATIONS

OF TOBE, START 1 WITH MRT

BY RACE AND SEX

Groups	TOBE/MRT	START 1/MRT
Black Male (N=34)	0.7603	0.6409
White Male (N-54)	0.7558	0.6629
Black Female (N-36)	0.6348	0.7063
White Female (N=71)	0.6662	0.4867

All correlations were significant at the .05 level of confidence.

Overall, males produced very consistent correlations when compared within and between racial groups and to the total study population. Females, on the other hand, showed marked differences within and between racial groups and compared to the total study population. Although the coefficients themselves varied considerably, the differences between correlations within and between racial and sexual groups were not significant at the .05 level.

In response to one of the primary concerns of this study, how well screening tests administered at the beginning of the kindergarten year identify pupils who will be identified as not ready for first grade by the MRT at the end of the kindergarten year, correlations were fairly high. As was previously stated, the correlations for the TOBE and the MRT and the START 1 and the MRT remained consistently high across individual racial - sexual groups as well as for the total study population. The only exception to this being the correlation for white females on the START 1 and MRT. These correlations justify consideration of the TOBE as part of a screening program at the beginning of kindergarten; however, basing remediation on these early scores should receive further attention, in that simply remediating weak or low areas on the TOBE, does not necessarily in itself increase or insure "readiness" for first grade at the end of the kindergarten year, using MRT scores as criterion.

The poor correlation between START 1 and START 2 revealed a lack of improvement during the kindergarten year of skills assessed by this instrument as well as the instruments sensitivity to developmental changes

in the individual. Subsequently, this in conjunction with the poor correlation between the START 1 and MRT limits its application as a predictor of unreadiness at the end of the kindergarten year or as a legitimate tool by which to develop programs and assign pupils to these programs at the beginning of the kindergarten year.

As for individual racial - sexual groups, the data again identified the TOBE's legitimate application in the early identification of potentially high risk students. Although the correlations themselves indicated a potentially greater applicability of the TOBE in the identification of males, the failure of the difference between male and female correlations to achieve significance at the .05 level would warn against such usage.
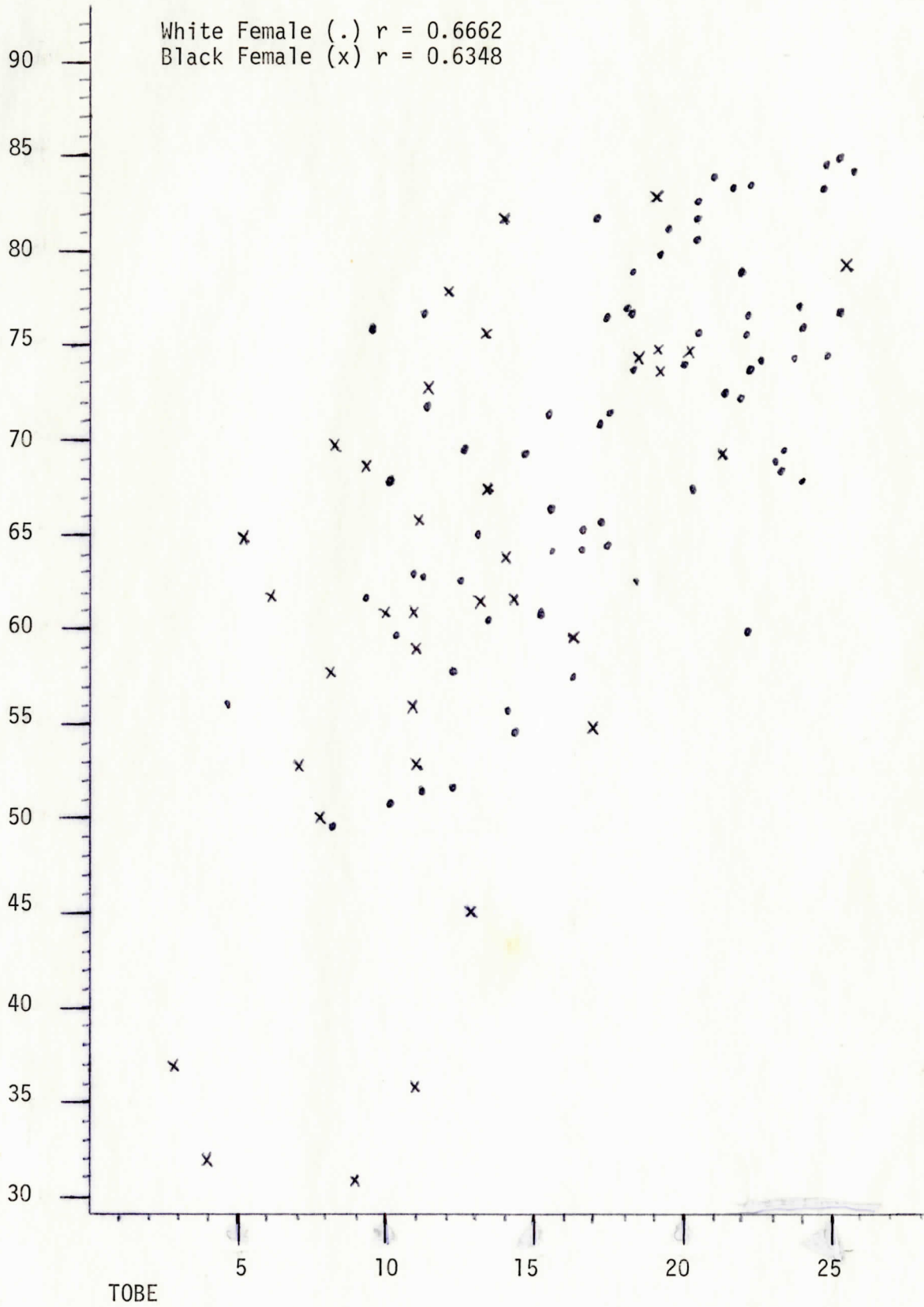
This study revealed interesting data regarding the instruments used and revealed several possibilities for additional research in the continually expanding program of preschool and kindergarten screening for the early identification of potentially high risk students. The TOBE could well become a reliable tool in the development of a relatively simple and efficient preschool screening program.

BIBLIOGRAPHY

Ahr, Edward A. The Development of a Group Preschool Screening Test of Early School Entrance Potentiality. Psychology in the Schools, 1967, 4 (1), 59-63.

Ahr, Edward A. Screening Test for the Assignment of Remedial Treatments. Chicago: Priority Innovations, Inc., 1968.

Atwell, Arthur A., Orpert, R. E. and Meyers, C. Edward. Kindergarten Behavior Ratings As A Predictor of Academic Achievement. Journal of School Psychology, 1967, 6, 43-46.

Bronfenbrenner, Urie. Is Early Intervention Effective? The Education Digest, February 1975, 17-20.

Correll, R. N. The Child Behavior Rating Scale. Beverly Hills, California: Western Psycholgical Services, 1962.

Culbertson, Frances M. and Gunn, Robert C. Comparison of the Bender Gestalt Test and Frosting Test In Several Clinical Groups of Children. Journal of Clinical Psychology, 1966, 22 (4), 439.

deHirsch, K., Jansky, J., & Langford, W. Predicting Reading Failure. New York: Harper & Row, 1966.

Dillard, Harry K. and Landsman, Myril. The Evanston Early Identification Scale: Prediction of School Problems From the Human Figure Drawings of Kindergarten Children. Journal of Clinical Psychology, 1968, 24 (2), 227-228.

Durkin, Dolores. What Does Research Say About the Time to Begin Reading Instruction? The Journal of Educational Research, October 1970, 64 (2), 52-56.

Faust, Margaret. Cognitive and Language Factors. In B. K. Keough (ed.), Early Identification of Children With Potential Learning Problems. Journal of Special Education, 1970, 4, 335-346.

Frostig, M., Maslow, P., Lefeuer, D. W., & Whittlesey, J. R. B. Developmental Test of Visual Perception. Palo Alto, Calif.: Consulting Psychologists Press, 1963.

Haring, Norris G. and Ridgway, Robert W. Early Identification of Children with Learning Disabilities. Exceptional Children, 1967, 33 (6), 387-395.

Hildreth, Gertrude H., Griffiths, Nellie L., and McGauvron, Mary E. Metropolitan Readiness Tests. New York: Harcourt, Brace & World, Inc., 1969.

Huberty, Carl J. and Swan, William W. Preschool Experience and First Grade Achievement. The Journal of Educational Research, March 1974, 67 (7), 311-316.

Katz, Lillian G. Perspectives on Early Childhood Education. The Education Digest, October 1973, 15-19.

Keogh, Barbara K. The Bender Gestalt As a Predictive and Diagnostic Test of Reading Performance. Journal of Consulting Psychology, 1965, 29, 83-84.

Keogh, Barbara K. School Achievement Associated With Successful Performance on the Bender Gestalt Test. Journal of School Psychology, 1965, 3, 37-40.

Keogh, Barbara K. and Becker, Lawrence D. Early Detection of Learning Problems: Questions, Cautions and Guidelines. Exceptional Children, September 1973, 5-11.

Keogh, Barbara K. and Smith, Carol E. Early Identification of Educationally High Potential and High Risk Children. Journal of School Psychology, 1970, 8 (4), 285-290.

Kirk, Winifred D. A Tentative Screening Procedure for Selecting Bright and Slow Children in Kindergarten. Exceptional Children, 1966, 33 (4), 235-241.

Koppitz, E. M. The Bender Gestalt Test for Young Children. New York: Greene and Stratton, 1964.

Mitchell, Blythe C. The Metropolitan Readiness Tests As Predictors of First Grade Achievement. Educational and Psychological Measurement, 1962, 22, 765-772.

Moss, Margaret H. Tests of Basic Experiences. Monterey, California: CTB/McGraw-Hill, 1970-71.

Parnell, Dale. Upgrading the Downgrader. Elementary Teachers. February, 1974, 6-9.

Rogalsky, Maryrose M. Screening Kindergarten Children: A Review and Recommendations. Journal of School Psychology, 1968-69, 7 (2), 18-27.

Spivack, George and Swift, Marshall. The Classroom Behavior of Children:
    A Critical Review of Teacher-Administered Rating Scales. The
    Journal of Special Education, 1974, 7 (1), 55-90.

Spodek, Bernard. Sources of Early-Childhood Curriculum. The Education
    Digest, January 1971, 49-52.

Tirrell, Richard I. Personal Communication, April 15, 1976.

Wang, Margaret C. The Accuracy of Teacher's Predictions on Children's
    Learning Performance. The Journal of Educational Research,
    July-August 1973, 66 (10), 462-465.

White Male (X) r = 0.7558
Black Male (.) r = 0.7603

22

APPENDIX A

TABLE 4   SCATTER DIAGRAM OF MALES' TOBE AND MRT RELATIONSHIP

White Female (.) r = 0.6662
Black Female (x) r = 0.6348

TOBE

APPENDIX B

TABLE 5 SCATTER DIAGRAM OF FEMALES' TOBE AND MRT RELATIONSHIP