THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE ESTABLISHMENT OF RELEVANT SIXTH-GRADE NORMS ON A READING TEST CONSTRUCTED FOR GRADES 7-9

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

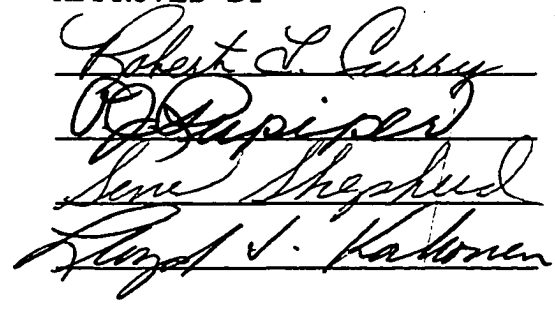
ZOE WALKER LEIMGRUEBLER

Norman, Oklahoma

1979

THE ESTABLISHMENT OF RELEVANT SIXTH-GRADE NORMS ON A READING TEST CONSTRUCTED FOR GRADES 7-9

APPROVED BY



DISSERTATION COMMITTEE

ACKNOWLEDGEMENTS

I wish to express appreciation to my committee chairman, Dr. Robert Curry, and to the members of the committee, Dr. Omer Rupiper, Dr. Gene Shepherd, and Dr. Lloyd Korhonen, for the kindness, help, and advice they provided unfailingly at needed points during the preparation of this study.

I wish to express deep appreciation to Dr. John Poarch who has consistently provided the kind of friendship and guidance which leads to both personal and professional growth.

I wish to thank Dr. Walter MacGinitie for granting permission to utilize his reading test and for encouraging me in this endeavor.

I wish to thank the administrators and teachers in Putnam City District without whose complete cooperation such a study could never have been conducted.

I wish to express appreciation to Mrs. Ruth McKillen for her patience, incentive, perseverance, and sharing of her artistic talents in typing a beautiful copy.

Above all I wish to express to my husband, Norm, and to our daughter, Robin, and to our sons, David, John, and Michael my deepest love and appreciation for the patience, understanding, loving support, and inspiration shown me during this endeavor.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF TABLES

THE ESTABLISHMENT OF RELEVANT SIXTH-GRADE NORMS ON A READING TEST CONSTRUCTED FOR GRADES 7-9

CHAPTER I

PRESENTATION OF THE PROBLEM

Introduction

Public school districts need to coordinate the efforts of teachers and counselors toward accurate placement of students in reading programs during the transition from elementary to secondary schools. Data utilized for placement purposes generally include (a) scores obtained from standardized, group-administered tests; (b) students' grade point averages; and (c) recommendations from elementary teachers. These data are particularly useful to school personnel when classes are arranged to facilitate efforts to meet students' academic needs in relation to varying ability levels, and/or when course or program options are offered (Brown, 1970, pp. 10-12).

In those schools which offer remedial, developmental, accelerated, or optional reading classes, appropriate placement of students is dependent to a great degree on certain

factors which relate specifically to the use of test data for this purpose:

1. The efficiency with which educators control those variables considered to be the "mechanics" of test usage is a crucial factor in accuracy of placement. Qualitatively speaking, this is dependent on the degree to which teachers and administrators (a) provide the most advantageous environmental conditions for testing the student; (b) utilize the most efficient test administration procedures; and (c) utilize the most efficient and accurate scoring and reporting techniques. Local school personnel must, by necessity, assume full responsibility for quality control in these areas.

2. The efficiency with which a specific standardized test measures and estimates the level of student performance in relation to the academic ability in question is an equally critical factor in accuracy of placement. Efficiency, in this sense, is dependent on (a) the degree of reliability and validity of the measuring instrument; and (b) the degree of relevancy of the test norms utilized to convert raw scores to appropriate transformed scores, such as percentiles or relative deviate scores. In the great majority of school districts, the behavior of teachers and administrators indicates their assumption that it is the test publishers' responsibility to control these factors, and that this responsibility is fulfilled when reliability coefficients and national norms are provided in the test manuals. Evidence of

such attitudes is found in Mehrens' and Lehmanns' statement (1973, p. 144) that "national norms are more commonly used than any other type of norms." While this practice is acceptable for those school populations which closely characterize the national norm group, it is highly questionable for those which do not.

Mehrens and Lehmann (1973, p. 141) state that "a norm group, to be appropriate, must be recent, representative, and relevant." Recency is important in that, if too great an amount of time has lapsed between the norms construction and their use at the local level, they can become outdated due to the rapid changes taking place in education, and/or the changing characteristics of the reference group. Brown (1970, p. 162) suggests that "norms should periodically be updated and old norms looked upon with appropriate skepticism." Representativeness is important in that, if the sample utilized for norming was inadequate in size, kind, or if sampling was done by convenience rather than by a particular sampling procedure, the norms may be biased and therefore inappropriate for use in any district except those which were utilized to form the norm group (Mehrens & Lehmann, 1973, p. 143-44). The relevance of the norms utilized is dependent upon "the degree to which the population sampled is comparable to the group with which users of the test wish to compare their students" (Mehrens & Lehmann, 1973, p. 144).

National norms developed for standardization of tests are based on the combined scores from school districts representing a wide variety of sizes, socioeconomic levels, and geographic areas. Therefore, the use of scales based on national norms may yield results which are unrealistic to a measurable degree when tests are administered in a school district whose population deviates widely from the national mean (Brown, 1970, p. 163). If scores from district test results at a given grade level indicate that the test is inappropriate for use at that level and/or with the norm scales provided by the publisher, it becomes advantageous to the district in question to (a) determine which of a series of leveled tests possess the most appropriate ceiling for the sample in question; and (b) to take the necessary measures to establish local norms on that particular test. The preparation of local norms thus "facilitates test score interpretation to the teacher, the parent, the students, and the community" (Mehrens & Lehmann, 1973, p. 146).

Need for the Study

During the Spring of 1974, a district-wide testing program was implemented in Putnam City School District, Oklahoma City, Oklahoma, in which the reading achievement levels of sixth-grade students were assessed. The goal was to provide an additional means by which, when used in conjunction with students' grade point averages, teachers' recommendations and achievement test scores obtained each year during

October, accurate placement of these students in seventh-grade reading programs could be achieved.

The instrument chosen for reading assessment was the Gates-MacGinitie Reading Test, Survey E, Forms 1, 2, and 3 (originally constructed and normed for use in grades 7-9; Teachers College Press, 1964). It had been determined through repeated use of the Gates-MacGinitie Reading Test, Survey D (for grades 4-6) with sixth-grade students during the previous six years that the ceiling and norms for this test were too low to provide an appropriate measure of reading achievement. While the resulting test data from Survey D could be used for relative ranking at the lower end of the range, the scores occurring at the upper end of the range appeared to overestimate reading achievement levels to such a degree that this test was viewed as inappropriate for placement purposes (e.g., comparisons of Survey D grade equivalent scores and students' actual instructional/operational reading levels at the upper end of the range yielded differences representing overestimates of 2.0 to 5.5 grade levels).

In an effort to find a more appropriate placement test, the Vocabulary and Comprehension subtests of the Gates-MacGinitie Reading Test, Survey E, Forms 1 and 2 were administered to sixth-grade students district-wide for three consecutive years (1974-76). The resulting data indicated that Survey E possessed a ceiling which was adequate for

this population. The district raw score means obtained on each of the parallel forms consistently fell within the range of raw score means obtained for the national norm group of seventh-, eighth-, and ninth-grade students, thus indicating that the upper limits of abilities possessed by Putnam City sixth-grade students appeared to be within the range of upper limits of ability measured by the test.

The resulting conclusion was that these students could neither be considered a representative sample of the original norm group for Survey D (grades 4-6) nor for Survey E (grades 7-9). This conclusion was drawn from a comparative analysis of exemplary test performance data gathered from the relevant groups. It was further concluded that Putnam City students represented a special norm group for which new and different norms were needed. Based on these conclusions, it was decided (a) to administer all subtests of Survey E, Forms 1, 2, and 3 to randomly sampled equivalent subgroups within and inclusive of the entire 1977 sixth-grade population (with the exception of students enrolled in classes for the Educable and Trainable Mentally Retarded); (b) to construct local norms from the resulting data; and (c) to utilize these norms to obtain the more appropriate scaled scores needed for placement of these students in seventh-grade reading programs.

The uniqueness of Putnam City students as a norm group and the resulting need for the present study (i.e., to provide for more accurate interpretation of obtained raw scores

on the Gates-MacGinitie Reading Test, Survey E) can be rationalized through a comparison of certain characteristics of this district and its population with those of the national norm group. These characteristics are related to size of the district, geographical location, socioeconomic level of the area served, academic aptitude of the students, and the general level of students' academic performance at any given grade level.

Putnam City is a suburban independent public school district which serves a 49 square-mile area in the northwest quadrant of Oklahoma City, Oklahoma. The school population includes 20,000 students in grades K-12 who attend 14 elementary schools, 4 junior high schools, and 3 high schools which are collectively staffed by 1,000 teachers and 40 administrative and/or supervisory personnel.

The socioeconomic level of the community ranges from lower-middle to lower-upper income, with the majority of families falling into the middle and upper-middle income brackets. The school population includes students of African descent (1.0%), Asian American (.5%), American Indian (5.8%), Spanish Surnamed Americans (.9%), and American Caucasian (91.7%) who are distributed somewhat evenly across the district.

A large number of Putnam City patrons exhibit an intense interest in the schools and in their children's academic progress. The Parent-Teacher Associations in 18 out of 21

of the local schools maintain yearly membership in excess of 100% (i.e., a parent/child ratio greater than 1:1; P.T.A. Annual Report, 1977). The level of interest is evidenced by the great number of parents found working as volunteers in every school in raising funds for printed materials, equipment, and media; serving as playground aides, library aides for book repair, boosters for the various sports activities; and serving as chaperones for field trips, sports activities, and social functions at the junior high and high school levels.

School attendance is maintained at 93% Average Daily Attendance at the elementary level, 92% at the junior high level, and 97% at the high school level, for a total weighted average of 94% for the district (1976-77 Attendance Report, Putnam City District Personnel Office). Attendance policies at all of the secondary schools (grades 7-12) require that parents contact the school office each day the student is absent to provide the reason. Upon returning to school, the student must obtain an office permit in order to be readmitted to class.

Of the 1,000 teachers in Putnam City District, 42% hold Masters' Degrees. This compares with a state average of 40% (Oklahoma State Department of Education, 1978). Of the number of district-employed teachers holding Masters' Degrees in any given semester, an average of 56 to 73 hold Standard Reading Specialists' Certificates (Office of Curriculum and

Instruction Surveys, 1973-77). All Reading Specialists in Putnam City Schools work as regular classroom teachers in (a) self-contained classes, language art blocks, or departmentalized reading situations at the elementary level; and in (b) departmentalized "Basic" English/reading situations at the secondary level. The percentage of reading teachers in the district who hold Reading Specialist Certification is 86 at the secondary level and 16 at the elementary level (these data represent averages over the 1973-77 school years).

The test results shown in Table 1 (California Test of Basic Skills, McGraw-Hill, 1970 edition) exemplify the academic performance of Putnam City students. These data represent the district means of grade equivalent scores for all students in the grade levels indicated (with the exception of those enrolled in classes for the Educable and Trainable Mentally Retarded). Several conclusions can be drawn from a comparison of these students' reading achievement scores with (a) their grade placement; and with (b) the levels at which they might be expected to achieve, based on a measure of academic aptitude:

1. As a group, their reading performance at a comparable grade placement is superior to that of the national norm group in 100% of the instances.

2. A measure of Academic Aptitude for this group is shown to be superior to that of the national norm group in 100% of the instances.

Table 1

Mean Test Scores by Grade Level

California Test of Basic Skills[a]--Reading Battery

Putnam City School District 1973-77

School Year	Grade Placement	Score	Reading Vocabulary	Reading Comprehension	Reading Total
1973-74	2.0	MGE[b]	2.5	2.4	2.5
		AGE[c]	2.3	2.2	2.3
		DIFF[d]	0.2	0.2	0.2
	4.0	MGE	4.6	4.6	4.6
		AGE	4.5	4.6	4.6
		DIFF	0.1	0.0	0.0
	6.0	MGE	6.7	6.6	6.7
		AGE	6.9	7.1	7.0
		DIFF	-0.2	-0.5	-0.3
1974-75	2.0	MGE	2.6	2.5	2.6
		AGE	2.3	2.2	2.3
		DIFF	0.3	0.3	0.3
	4.0	MGE	4.7	4.8	4.7
		AGE	4.7	4.7	4.7
		DIFF	0.0	0.1	0.0
	6.0	MGE	7.5	7.9	7.6
		AGE	7.2	7.8	7.5
		DIFF	0.3	0.1	0.1
	8.0	MGE	10.1	10.2	10.1
		AGE	9.4	9.6	9.5
		DIFF	0.7	0.6	0.6
1976-77	2.0	MGE	3.0	2.9	3.0
		AGE	2.5	2.4	2.5
		DIFF	0.5	0.5	0.5
	4.0	MGE	5.2	5.5	5.4
		AGE	5.0	5.2	5.1
		DIFF	0.2	0.3	0.3
	6.0	MGE	7.7	8.0	7.7
		AGE	7.5	8.0	7.7
		DIFF	0.2	0.0	0.0
	8.0	MGE	10.4	10.3	10.4
		AGE	9.9	10.1	10.0
		DIFF	0.5	0.2	0.4

[a]McGraw-Hill, 1970 Edition.

[b]MGE: Mean Grade Equivalent Score for entire grade tested.

[c]AGE: Antitipated Grade Equivalent for entire grade tested: an expectancy of achievement based on students' performance on the Short Form of Academic Aptitude (McGraw-Hill, 1970 Ed.).

[d]DIFF: Difference in the MGE and AGE.

3. In 73% of the instances, these students' Reading Achievement levels are shown to exceed the levels at which they can be expected to achieve (based on their Academic Aptitude).

4. In 18% of the instances, these students' Reading Achievement levels are shown to be equivalent to their Anticipated Grade Achievement levels.

5. Only 9% of the instances show students' Reading Achievement levels to be below their Anticipated Grade Achievement levels.

6. For each of the school years represented, Putnam City students appeared to have fewer characteristics in common with the national norm group with each successive grade level tested. That is, the difference in the Grade Placement (mean for the national norm group tested at that grade level) and the actual Mean Grade Equivalent became greater as students advanced in the grade levels.

The use of national norms with students such as those attending Putnam City Schools leads to a somewhat distorted view of achievement levels. More relevant norms are needed to help determine the number and kind of reading classes at the seventh-grade level and to place students appropriately.

Purpose of the Study

The Gates-MacGinitie Reading Test, Survey E, Forms 1, 2, and 3 appeared to be an appropriate measurement instrument

for use with sixth-grade students in Putnam City Schools. The purpose of this study was to bring relevance to its use as a placement tool through establishing norm scales from which meaningful standard and percentile scores could be derived.

Statement of the Problem

The problem was to establish local/special-study norms for Putnam City School District on Survey E, Forms 1, 2, and 3 of the Gates-MacGinitie Reading Test. Norms were established at the sixth-grade level using the total district population for that grade (with the exception of students enrolled in classes for the Educable and Trainable Mentally Retarded). This test was written and nationally normed in 1964 for students in grades 7-9. Each form contains Speed and Accuracy, Vocabulary, and Comprehension subtests for which separate norms were established.

Two related problems were explored in which the following questions were considered in regard to each of the subtests:

1. Was there a statistically significant difference in the raw score mean obtained for Putnam City students tested at grade placement 6.5 and the raw score equivalent of the interpolated grade equivalent score of 6.5 on the original norm scale?

2. Would a comparison of the Putnam City and Metropolitan School Study Council Norms established for Survey E at

grade 6.5 (Gates and MacGinitie, 1965) confirm the relevance of MSSC Norms for Putnam City use?

Hypotheses

The following hypotheses were tested at the .05 level of significance:

Ho_1 There are no statistically significant differences among the IQ score means obtained on the Short Form Test of Academic Aptitude for each of the samples represented by Forms 1, 2, and 3 of the Gates-MacGinitie Reading Test, Survey E.

Ho_2 There are no statistically significant differences among the raw score means for Number Correct on the Speed and Accuracy subtest for the three samples represented by Forms 1, 2, and 3 comprising the Putnam City norm group.

Ho_3 There are no statistically significant differences among the raw score means obtained on the Vocabularly subtest for the three samples represented by Forms 1, 2, and 3 comprising the Putnam City norm group.

Ho_4 There are no statistically significant differences among the raw score means obtained on the Comprehension subtest for the samples represented by Forms 1, 2, and 3 comprising the Putnam City norm group.

Ho_5 There are no statistically significant differences in the Speed and Accuracy raw score means (Number Correct) obtained on any of Forms 1, 2, or 3 for the Putnam City samples and the comparable subtest means obtained for students comprising the seventh-grade portion of the original norm group.

Ho_6 There are no statistically significant differences in the Vocabulary raw score means obtained on any of Forms 1, 2, or 3 for the Putnam City samples and the comparable subtest means obtained for students comprising the seventh-grade portion of the original norm group.

Ho_7 There are no statistically significant differences in the Comprehension raw score means obtained on any of Forms 1, 2, or 3 for the Putnam City samples and the comparable subtest means obtained for students comprising the seventh-grade portion of the original norm group.

Ho_8 There are no statistically significant differences in the Speed and Accuracy raw score means (Number Correct) obtained for Putnam City District's samples tested at grade placement 6.5 and the raw score equivalents of the interpolated grade equivalent scores for 6.5 on the original norm scales for any of Forms 1, 2, or 3.

Ho_9 There are no statistically significant differences in the Vocabulary subtest raw score means obtained for Putnam City District's samples tested at grade placement 6.5 and the raw score equivalents of the interpolated grade equivalent scores for 6.5 on the original norm scales for any of Forms 1, 2, or 3.

Ho_{10} There are no statistically significant differences in the Comprehension subtest raw score means obtained for Putnam City District's samples tested at grade placement 6.5 and the raw score equivalents of the interpolated grade equivalent scores for 6.5 on the original norm scales for any of Forms 1, 2, or 3.

Limitations of the Study

The test scores with which local norms were constructed in this study were gathered from the 1976-77 population of sixth-grade students in Putnam City School District, Oklahoma City, Oklahoma. With the exception of classes for the Educable and Trainable Mentally Retarded, all students were tested, and all scores were used. While the resulting norm scales may be used with successive Putnam City sixth-grade students, these data should not be generalized to populations of other school districts.

CHAPTER II

A REVIEW OF THE LITERATURE

Introduction

Norm referenced measures are administered for the purpose of obtaining performance data which indicate a subject's or group's relative ranking in comparison with specified others on the same measure (Schrader, 1958, p. 922; Angoff, 1971, p. 534). These normative data are subsequently utilized for decision-making at the institutional level (e.g., selection, classification, placement, educational or vocational guidance, public relations, or other) and/or at the individual level (e.g., vocational choice, educational choice, and other personal decisions one makes about himself) (Schwartz, 1974, p. 2). For such decisions to be accurate, the data base on which they are made must be not only statistically valid, but relevant (Waggener, Starr, & Hopkins, April 1974, p. 1; Seashore & Ricks, May 1950, p. 16).

Mehrens and Lehman (1973, p. 144) and Angoff (1971, p. 536) indicated that, in regard to decision-making at the local level, the relevancy of normative data is dependent on

the degree to which the original norm group and the sample being tested are comparable on certain essential variables. At the individual level, these variables include educational level, age, sex, grade placement, race, aptitude, socioeconomic level of parents, geographic location of residence, attitude (level of motivation at the time of testing), and, for vocational guidance purposes, the individual's intended field of study and/or career interests.

Those essential variables related to institutional use of normative data include size of school, geographic location, type of support (public, parochial, or independent private), pupil/teacher ratio, per/pupil expenditures, curricular emphasis, and the proportion of students who are college-bound. Certain community-based variables which affect institutional decisions include the region in which the community is located (e.g., South vs. North), type of community (rural, urban, suburban, size of geographical area served, and density of population), socioeconomic level of the community, and the presence/absence and size of the community library.

Two major points of emphasis regarding the relationship of these variables to test norming were found throughout the literature. The American Psychological Association's Committee on Test Standards (Supplement to the Psychological Bulletin, March 1954, p. 36), Mehrens and Lehmann (1974, pp. 143, 146-47), Angoff (1971, p. 551), and others strongly recommended that, in order to preserve the accuracy and

subsequent usefulness of the norms established, the researcher should (a) maintain strict control of essential variables in designing the sampling procedures (preferably through stratified random sampling where stratification is done on all relevant independent variables); and (b) provide accurate, complete descriptions of samples, populations from which they were drawn, variables considered in the sampling procedures, and populations to which the norm scales can be generalized.

Major test publishers, in preparing national norms for general use, attempt to utilize large, randomly sampled groups which, when combined, represent a cross-section of the population in relation to a given set of variables. Mehrens and Lehmann (1973, p. 144) rated the sampling methods employed by most major test publishers who report national norms as "reasonably satisfactory." Schrader (1958, p. 923) pointed out that, from the user's viewpoint, national norms have the distinct advantage of being simple, definite, and unique. Perhaps the greatest advantage of all is their availability to the test user--a condition which at least partially contributes to the fact, as Mehrens and Lehmann stated in 1973 (p. 145), that "national norms are more commonly used than any other type of norms."

The disadvantages in using national norms have their source in errors of measurement and sampling error (Mehrens & Lehmann, 1973, p. 143). With great care and skillful attention to test construction, strict control and testing

procedures, and adequate statistical treatment, errors of measurement can be detected and compensated for in the norming and reporting processes. Sampling error, however, is neither so easily detected nor remedied. Too often, sampling must be done by convenience rather than by design due to the unwillingness of school administrators to co-operate in norm studies. That is, (a) entire school populations are often tested rather than a randomly selected sample of students; and (b) often the administrators of those schools needed to construct a "national average" profile are not the ones who will allow their students to be tested. This results in sample bias, and merely increasing the sample size does not solve the problem (Mehrens & Lehmann, 1973, p. 143).

Mehrens and Lehmann (1973, p. 145) emphasized that, among the various limitations in the use of national norms, the most serious of all is that "often they simply do not provide the comparison we need." To combat this inadequacy, the authorities advocated the development and use of highly specialized norms, such as local, differentiated, school mean, and special group norms. The potential for subsequent usefulness inherent in establishing such norms increases in direct proportion to the magnitude with which the local sample to be tested deviates from the national norm group (i.e., the greater the deviation on a given set of variables, the more relevant are the specialized norms established).

Related Special Norm Studies

Examples of special norm studies are cited in the following pages. The major point of similarity is the underlying need and resulting purpose of each--i.e., to provide more relevant norm scales with which to interpret raw scores on a given test or set of tests for specified groups in question. Each case cited therefore represents an effort to renorm one or more instruments previously normed for more general use.

The major point at which these studies diverge is the intended level of use of the resulting norm scales. This point will therefore serve as a basis on which to categorize the studies discussed, beginning with those involving large, diverse samples from which, according to the authors, the resulting norm scales can be generalized to similar specialized groups across the nation.

National Level

One special-purpose norm study was conducted through the combined efforts of the U.S. Armed Forces Institute and the Test Department of Harcourt Brace Jovanovich, Inc. (Waggener et al., 1974). Waggener, a military educational counselor, sought to insure the relevance of the test scores utilized to make educational decisions for specific groups within the military population.

The Adult Basic Learning Examination (ABLE) was normed and validated in terms of its utility in accurately predicting

success on the General Educational Development (GED) tests. ABLE is an adult achievement battery written for three levels--I, II, and III, for Grades 1-4, 5-8, and 9-12, respectively. Each level contains four subtests--Vocabulary, Reading, Spelling, and Arithmetic (Computation and Problem Solving).

Sampling methods included the selection of 1,990 recruits from seven Army, Navy, Air Force, and Marine Corps training stations, and 1,383 GED candidates from six Army bases. All were administered the SelectABLE, a short screening test, and the appropriate level of ABLE, with one exception: none of the GED candidates were administered Level I of the ABLE. In addition, all recruits were administered the Armed Forces Qualification Test (AFQT) prior to induction into the Armed Forces. A survey to obtain demographic data was completed by each.

Using the total military sample, ABLE was normed by subtest at each level. The norm scales established were expressed in percentile ranks and stanines. Means, medians, quartiles, and standard deviations were generated for the ABLE and SelectABLE by educational level for each of the four services (Waggener et al., 1974, p. 2).

Intercorrelations were generated by level among the ABLE subtests. In addition, correlations of the ABLE subtests at each of the three levels were obtained with the AFQT,

GED subtest scores, GED totals, SelectABLE scores, and with various military Classification Battery scores.

Using the ABLE scores as predictors, a multiple regression analysis was calculated for each of the GED subtests. Prediction of GED scores from ABLE performance was thus made possible through the resulting regression equation.

With the aid of these new and more valid data, military counselors were able to make more accurate judgments regarding recruits' levels of performance in the basic skills. Through use of the predictive capabilities developed, they were able to initiate prescriptive programs with greater confidence and greater potential for success.

Prior to completion of this project, each of the military services traditionally developed its own vocational aptitude battery for use in the selection and classification of potential enlistees. However, in 1974, the Department of Defense (DOD) directed the military services "to jointly develop and employ a single aptitude battery for use in both the High School Testing Program and the Armed Forces Examining and Entrance Stations (AFEES)" (Armed Services Vocational Aptitude Battery Counselor's Guide No. DOD1304.12X, pp. 8, 30). The purpose was to standardize for all services both the battery and testing practices used (a) to determine mental qualifications for enlistment; and (b) to provide the information on aptitudes needed by the various branches of the Armed Forces for initial classification and assignment

decisions. The need for such uses necessitated special norming procedures which, in their uniqueness, were of particular relevance to the present study.

The battery developed in the DOD effort, the Armed Services Vocational Aptitude Battery (ASVAB, Forms 5, 6, and 7) contains the following subtests:

1. Speed Tests: Numerical Operations and Attention to Detail.

2. Power Tests: General Information, Word Knowledge, Arithmetic Reasoning, Space Perception, Mathematics Knowledge, Electronics Information, Mechanical Comprehension, General Science, Shop Information, and Automotive Information.

The samples utilized were intended to be representative of the national high school population (Adkins, October 1976, p. 5). Selection of schools was made from the Office of Civil Rights (OCR) Directory of Schools (Department of Health, Education, and Welfare, 1972), which, in itself, represented 91.8% of the enrolled pupils in the United States. The variables utilized for stratification included ten geographic regions, as determined by the first digit of the zip code (0-9); size of the student body in grades 9-12 (0-500 students, 501-1,000 students, and over 1,000 students); and percent of black enrollment (0-20, 21-40, 41-60, 61-80, and 81-100) (Adkins, October 1976, p. 5). The selection process, consisting of two stages, yielded a sample of 17,934 males and 17,357 females for a total sample size of 35,291.

Percentile norms for individual ASVAB-5 subtests were established by grade and by sex within each grade. The sample used for developing these norms included 35,291 high school students tested during the 1975-76 school year.

The special set of norms developed to meet vocational/educational counseling needs were based on six specific ability "composites"--Verbal, Analytical/Quantitative, Clerical, Mechanical, Trade Technical, and Academic Ability. The first five "composites" listed were actually the five major factors obtained in factor-analyzing a correlation matrix of the ASVAB-5 subtest results. The sixth composite, Academic Ability, was composed of the Word Knowledge and Arithmetic Reasoning tests. This composite was added because of its value in high school counseling as a predictor of educational success.

The six new "composites" were normed on a stratified random sample of 8,000 subjects selected from a population of 900,000 ninth- to twelfth-grade students tested with the new ASVAB-5 during the school year 1976-77. Percentile norms were reported by total grade level and separately by sex within each grade for each of the six "composites." The School Testing Program utilizing the ASVAB-5 was begun in July 1976. The parallel forms, ASVAB-6 and ASVAB-7 became the official selection and classification tests for all prospective armed services recruits beginning in January 1976.

Forms 6 and 7 are administered only at the Armed Forces Examining and Entrance Stations.

Another large-scale effort to produce more relevant norms was the Anchor Test Study (ATS), conducted by the Educational Testing Service and financed by the U.S. Department of Health, Education, and Welfare (Loret, Seder, Bianchini, & Vale, 1975). The purpose of this study was to increase the effectiveness of national and state educational assessment programs through providing a suitable method of comparing test scores obtained on the most commonly used reading tests.

In all, eight tests were administered at the fourth-, fifth-, and sixth-grade levels. Included were the California Achievement Tests (1970 ed., CTB/McGraw-Hill); Comprehensive Tests of Basic Skills (1968 ed., CTB/McGraw-Hill); Gates-MacGinitie Reading Tests (1964 ed., Teachers College Press); Iowa Tests of Basic Skills (1971 ed., Houghton-Mifflin); Metropolitan Achievement Tests (1970 ed., Harcourt Brace Jovanovich, Inc.); Sequential Tests of Educational Progress, STEP, Series II (1969 ed., Educational Testing Service); SRA Achievement Series (1971 ed., Science Research Associates); and the Stanford Achievement Tests (1964 ed., Harcourt Brace Jovanovich, Inc.).

Stratified random sampling was utilized in both the 1972 and 1973 data collections. The 918 schools selected were considered to provide proper representation in regard to type

of school (public, Catholic, and private non-Catholic); size (district or county); percentage of minority children in the school or community; income level of the immediate community; geographic region; and degree of community urbanization (1970 census) (Loret et al., 1975, p. 7). A total of 150,000 students were tested--50,000 at each of the three grade levels. Each test administered was paired with its alternate form to obtain parallel-form reliability estimates.

The Anchor Test Study efforts yielded the following norm scales and equivalency tables: (a) raw score equivalency tables (with which an individual student's raw score on any one test can be converted to an equivalent raw score on any of the remaining tests); (b) individual score norms tables (with which a student's raw score on any given test can be converted to equivalent percentiles and stanines on any other of the seven remaining tests); (c) school mean norms tables (with which a school's mean on any given test can be translated into national percentile ranks and/or stanines). Loret et al. (1975) indicated that

> The size of the school sample and degree of cooperation obtained in the study were of such magnitude that the resulting school-mean norms provide reliable information . . . about the relative level of reading performance for individual schools within or across districts or states (p. 5);

and (d) tables of individual score norms (with which a student's percentile ranks from the test publisher's norms and those of the Anchor Test Study can be compared).

State, District, and Local

Using the Anchor Test Study results as an aid, the Washington State Department of Education developed a statewide achievement profile based on total reading scores at fourth-, fifth-, and sixth-grade levels (Rasp & Stiles, April 1976 and Rasp, December 1976). Rather than mandate a specific test to be administered statewide, Rasp and his associates collected data on those tests included in the ATS which were already in use in Washington classrooms. Data on six of the eight tests were usable.

The districts were divided into ten categories based on size of school population (e.g., 0-299, 300-499, 500-699, etc.). An attempt to sample 20% of the districts in each category with 10% replacement yielded a range of 6.5% to 11.8% in the sample (Rasp & Stiles, April 1976, p. 3).

Total reading score means and standard deviations were reported for the combined districts in each size category. Means and standard deviations at the state level were also reported for each test, the former of which appears to have been obtained without the aid of statistical weighting procedures to equally distribute the effects of district size.

The test results indicated that those schools whose students achieved the highest levels in reading were located in districts of 2,000 to 2,999 enrollment, and the lowest-achieving schools were located in districts enrolling 300-499 students. By using the ATS school mean conversion tables, a

building principal could therefore determine the relative standing of his students in comparison with that of the entire district, the state, and with the median scores of the nationwide ATS samples.

Rasp and his colleagues (December 1976) reported that there were serious obstacles to be overcome in the state level use of ATS norms. These obstacles, viewed by the investigators as surmountable through more efficient planning, organization, and improved computer programming, were listed as follows: (a) insufficient sampling percentages within the district size categories; (b) gathering of raw data in many diverse, inequitable forms (leading to problems in data interpretation and transformation before it could be fed into the computer); (c) lack of ease and efficiency with which the conversion of raw data to ATS equivalents was made; (d) failure to regulate specific test forms utilized at the classroom level with those employed in the Anchor Test Study; and (e) the projected use of tests which either have been revised since the ATS was completed or which might be revised in the future. In summary, Rasp discussed these problems, recommended specific solutions, and stated that he would consider the use of ATS norms to be feasible at the state and district levels, provided that proper precautions were taken (December 1976, p. 8).

Huyser, Fisher, and Olsen (April 1973), in reporting the Michigan Educational Assessment Program (MEAP) results, described a quite different approach to state level norming

than that utilized in the Washington State Department of Education project. Assisted by the Educational Testing Service, the Michigan State Department of Education conducted the MEAP in order to provide individual schools and districts with normative needs assessment data to use as a basis for making decisions regarding (a) designs for improving educational programs; and (b) allocation of available resources.

District and school data were gathered for a total of 14 specific measures in the following categories: (a) human resources; (b) ethnic background of students; (c) size of district/school; and (d) achievement at grades 4 and 7 in Reading, Word Relationships, Mechanics of Written English, Mathematics, and a composite of basic skills.

Cluster sampling for norm construction was stratified on the basis of the major community type served (Huyser, et al., 1973, p. 31). The categories for community types included (a) Metropolitan Core Cities; (b) Cities (10,000 or more in population; (c) Towns (2,500-9,999 population or rural communities impacted by large military installations); (d) Urban Fringe; and (e) Rural (populations of less than 2,500 or RFD route). Of 604 districts in existence at the time (1972), only the 529 which were organized to operate K-12 programs were utilized in constructing the norms. Results on the remaining 75 districts were reported separately.

A norm-referenced test was developed especially for the MEAP by the Educational Testing Service. This instrument,

administered to all students in grades 4 and 7, was designed to measure Word Relationships, Reading Mechanics of Written English, and Mathematics. A Basic Skills Composite Achievement score was also computed for each student in grades 4 and 7.

District, school, and pupil norm tables were constructed separately at grades 4 and 7 for both statewide and community-type stratified distributions. Appropriate norming data and test results were distributed to each district and local school. This information, along with data reflecting the measurement categories, was accompanied by a manual which provided an explanation of the assessment procedure and results (Huyser et al., April 1973). Local school and district personnel were encouraged to utilize the results charts provided by the Michigan State Department staff in drawing profiles to emphasize need areas.

Special-Study Norms

Angoff (1971, p. 540) stated that "norms are useful to the extent that the reference group is meaningful to the user." Special-study norms, according to Angoff (1971),

> Capitalize on the familiarity to test users of certain well-known groups of students and, in a manner of speaking, yield as much information about the sensitivity of the test and its ability to differentiate within both high- and low-scoring groups of students as it does about the groups themselves. (pp. 540-41)

The following discussion identifies and describes four examples of special-study norms construction. In each case,

new and more relevant norms were sought for well-known groups whose special characteristics were thought to be a deterrent to the use of existing norms.

Pearson (1962, p. 10) conducted a study in which the primary purpose was to establish group intelligence norms for blind children in fourth, fifth, and sixth grades enrolled in state schools for the blind in Oklahoma, Arkansas, Tennessee, Kentucky, Missouri, Kansas, and Nebraska. Simple cluster sampling was utilized, yielding a sample size of 236. Of the 236 students tested, only those scores from the 197 white students were used for norms construction.

The School and College Ability Test (SCAT), Level 5, Forms A and B were transcribed to Braille and Large Type for use with blind and partially-sighted students. Special instructions and answer sheets, appropriately adapted to the needs of these students were used. The general directions for administering the SCAT were followed explicitly with the exception of time limit. Each student was allowed to complete all test items.

Means and standard deviations were reported separately by grade level for both the oral and written administrations of the Braille and Large Type forms of the test. The norm scales were expressed in terms of converted score intervals and their corresponding frequencies, cumulative frequencies, and percentile ranks. Comparisons were made of students' SCAT converted scores with their verbal and performance IQ

scores on the Wechsler Intelligence Scale for Children (WISC) in all cases for which these scores were available.

The Pearson study demonstrated that certain aptitude tests designed for group administration with sighted students can be transcribed and used effectively with blind and/or partially-sighted subjects. The procedure utilized and the resulting norm scales thus provided a relevant and useful tool for estimating abilities among this special group of students.

Another study in which special group norms were established was conducted by the Arizona State Department of Education (Evard & McGrady, 1974) as a part of a larger study designed to identify the percentage of school children possessing handicaps. Known as the Arizona Prevalence Study (Evard & McGrady, 1974, p. 3), the larger project involved screening school children for mental retardation, learning disability, communication disorder, deafness, blindness, or a combination thereof.

The substudy, conducted by Evard and McGrady, concerned those children identified as having communication disorders. In determining the criteria for identification, the authors questioned the relevance and validity of using a test designed for one population to measure the magnitude of disorder among subjects from a strikingly different ethnic background. Having determined that existing norms were not relevant, Evard and McGrady established norms for (a) the

entire population of communication-disordered students in Arizona and (b) various ethnic subgroups within that population.

The instrument utilized was the Illinois Test of Psycholinguistic Abilities (ITPA). Norms were constructed for two subtests of the ITPA: Auditory Association and Grammatical Closure.

The population sampled consisted of Papago Indians, Mexican-Americans, Blacks, and Anglo-Americans in grades 1, 3, 5, 7, and 9. The sample of 976 children represented a random sampling, stratified on the basis of ethnicity, urban-rural residence, sex, and age. Due to residence location conditions existing at the time, the sample represented both urban and rural Mexican-Americans and Anglos, rural Papago Indians, and urban Blacks. Half of the subjects were from non-English-speaking homes (Evard & McGrady, 1974, p. 4).

The resulting norm scales, expressed in means and standard deviations by grade, sex, and subgroup, were compared with existing ITPA norms. The investigators concluded that (a) "the application of geographic norms tends to reduce the number of children who would be considered 'disordered' as compared to using the ITPA norms" (1974, p. 5); (b) "the use of special group norms in this and similar instances is crucial to accurate educational placement" (1974, p. 6); and that (c) "a test is valid for a specific group only if relevant norms have been developed for that group" (1974, p. 1).

While Evard and McGrady (1974) emerged from their study convinced of the necessity of special norms for each "special group," other investigators have concluded that, for their purposes, this practice is not always necessary. Broadley and Broadley (1975) focused on providing relevant local norms on the Burt-Vernon Graded Word Reading Test for eight rural school districts in the County of Newell No. 4, Alberta, Canada. The purpose of this study was to compare rural reading attainment results with urban norms established during Vernon's 1973 Calgary Restandardization Study (Vernon, 1973, p. 72). The Calgary norms had been utilized from 1973 to 1975 in the rural county schools.

Drawing from a total county school population of 1,500, the sampling was done through the selection of every seventh name from each of the eight school district rosters for grades 1-9. This yielded a sample size of 170, the subjects of which were distributed equally by grade level and sex.

Replicating Vernon's 1973 restandardization of the Burt-Vernon Word Graded Reading Test, the Broadleys utilized norming methods which employed a scattergram, drawn to show age in years and months on the x axis and test scores on the y axis. The median age of children who obtained each score was computed. A smoothed curve drawn through the median ages of each group of ten scores (0-9, 10-19, 20-29, etc.) proved to be identical. A comparison of Vernon's 1973 Calgary norms with the Albertan Rural Norms upheld the hypothesis of little

difference (.05) for reading attainments above the median age 7.5. A significant difference was found in comparing mean scores at each median age from 6.7 (the lowest in the rural sample) through 7.5, t (16) = 3.76, $p < 0.01$, thus lending support to the hypothesis that Vernon's downward prorated norms, necessitated by later sampling, would not reflect the actual learning curve (Broadley & Broadley, 1975, p. 3). The curve appears to demonstrate that the rural students attain later, but more rapid development in the acquisition of a sight word vocabulary than do the urban students. However, beyond the median age 7.5, word recognition attainment rates appear to be equivalent between the two groups. The results (a) indicated a prorating of Vernon's 1973 norms too far downward; and (b) confirmed that the Burt-Vernon Graded Word Reading Test could be generalized to the rural Albertan population.

The final study examined in these pages, Metropolitan School Study Council (MSSC) Norms for the Gates-MacGinitie Reading Tests (Gates & MacGinitie, 1966), is most relevant for the present study in its intended purpose and therefore exemplifies the need for local norms. Having developed the original national norms for these tests in 1964, Gates and MacGinitie renormed all forms of the tests in order to provide "more meaningful and useful normative data for schools whose pupils were typically above average in ability and aspirations" (Gates & MacGinitie, 1966, p. 2).

The Metropolitan School Study Council constitutes a group of school districts and systems located in suburban communities near New York City. At the time of testing (1965), the MSSC was comprised of 70 school systems, most of which represented communities which "house a large proportion of business and professional people" and which had "long been concerned with providing excellent educational opportunities for their children" (Gates & MacGinitie, 1966, p. 3). The population of the MSSC districts performed at "a little more than half a standard deviation above the national average on standardized aptitude and achievement measures" (p. 3). Average median IQ scores were 118 in a 1965 testing (Lorge-Thorndike Intelligence Test, 1957 ed.) and 109 in an administration of the Lorge-Thorndike Intelligence Test (1964 Multilevel Edition), completed in connection with the MSSC study.

In addition, the investigators stated that education was taken seriously in the MSSC schools at the time, as evidenced by a mean holding power of 93.1% as compared with a nationwide mean of 81.5%. (Standard deviations were 3.8 and 11.6 for the two means, respectively.) Of the high school graduates, 60% later attended colleges or junior colleges, and "17% went on to other post-high school education" (Gates & MacGinitie, 1966, p. 4).

Participation by schools in the 1965 norms study was placed on a voluntary basis. The 57 elementary, junior high, and high schools that chose to participate represented nearly

half of the total number of MSSC schools. The study involved approximately 12,700 students in grades 1-9.

The sampling procedure utilized required an initial grouping of the 57 participating schools into two groups based on data gathered on socioeconomic status and past academic achievement. The odd-numbered grades were tested in one group and the even-numbered ones in the other group. Cluster sampling methods were utilized to the extent that, for any given grade level tested in a particular school, all classes in that grade were tested.

In order to provide an additional control variable, Verbal IQ scores were gathered from a 1965 administration of the Lorge-Thorndike Intelligence Tests (1975 ed.) in 14 schools, and a later administration of the 1964 Multi-Level Edition (Lorge-Thorndike) in an additional 34 schools. The Multi-Level Edition was administered in grades 4, 5, 7, 8, and/or 9, depending on whether a school fell into the "odd" or "even" grouping. No intelligence scores were available for nine of the 57 participating schools.

A frequency distribution of IQ scores was constructed for each Lorge-Thorndike edition, using the total sample as the data base and the classroom as the unit. A final selection of subsample schools was then made in which the frequency distribution of scores across grade levels was constructed to conform as nearly as possible to the two larger distributions. Resulting subsamples used in norming

the Gates-MacGinitie Tests ranged in number from 600 at the first and second grades to 1,500 at the eighth grade level.

Gates and MacGinitie (1966) reported that representativeness of the participating schools and districts

> was checked on the basis of 1965-66 figures for per/pupil expenditure, professional staff ratio, tax rate on assessed property valuation, and teachers' salary median . . . and no adjustment of the total sample appeared to be necessary (p. 10).

Norms, expressed in standard scores and percentiles, were constructed for all Gates-MacGinitie Tests in existence at the time (Surveys A-E). Since the population of students in MSSC schools had previously exhibited abilities and aptitudes measuring well above those of the national norm group, the investigators were able to norm Surveys B-E at 0.3 to 2.0 grade levels below the levels of usage established in the original norming. From a single October testing, norm scales were constructed for use in February and May administrations of equivalent test forms for all newly established MSSC levels. Thus the test-retest pattern established in the original 1964 norming was maintained in the MSSC norms construction.

The MSSC project was especially relevant to the present study in that, like the New York students in question, the Putnam City students also perform well above the national average. In order to provide a reading test "having the best measurement characteristics" (Gates & MacGinitie, 1965, p. 9) for the local population of sixth-grade students, it was

necessary to establish sixth-grade norms for Survey E, the level originally written and normed for grades 7-9. While the performance and abilities in MSSC and Putnam City Districts appeared strikingly similar, there were two reasons to suspect that the local use of MSSC norms might not be appropriate:

1. Since the MSSC norms were constructed in 1965, the time lapse of 12 to 13 years could adversely affect their usefulness in Putnam City District due to changing emphases in skill instruction, changing characteristics of the respective norm groups, and general changes in the society.

2. The comparison of academic aptitude scores (normally a reliable variable with which to compare groups) could not be depended upon in this instance because of (a) the differences in standardized tests used, and (b) the time lapse of 13 years between the respective test administrations. Either of these variables could negate a comparison of norms.

Gates and MacGinitie (1966, p. 3) stated that the median average Verbal IQ score for the 14 schools tested in 1965 (Lorge-Thorndike, 1957 ed.) was 118, and a corresponding measure for an additional 34 schools tested during the MSSC study (Lorge-Thorndike, 1964 Multi-Level Edition) yielded a median average of 109. (The difference of nine IQ score points in these two measures was attributed to differences in the two Lorge-Thorndike editions utilized.)

In the Putnam City norm group, the mean IQ scores were 105.61, 105.94, and 105.93 for the three samples ($\bar{X}_n$ = 493) represented by the Gates-MacGinitie Reading Test, Survey E, Forms 1, 2, and 3, respectively. The median IQ for each of the three samples was 106.

While both the MSSC and Putnam City norm groups exhibited above-average abilities based on national norms for the respective tests administered, there was not sufficient evidence of norm group equivalence in this fact to justify local use of the MSSC norms without a preliminary comparison of their respective reading score distributions. This comparison is discussed and illustrated in Chapter IV of the present study.

CHAPTER III

METHODS AND PROCEDURES

Sample Selection

The most common method of sampling utilized in norms construction is simple cluster sampling. In this method, "all students in each school selected are tested and included in the norms, provided they are at the proper grade levels or level" (Lord, 1959, p. 249). For the test publisher who must establish national norms, cluster sampling is simpler, less expensive, and much more readily accepted by school administrators than the method in which the test is only administered to individuals who are randomly sampled from selected schools. However, the use of cluster sampling in constructing national norms has been highly criticized (Schrader, 1958, p. 924; Lord, 1959, p. 249; and Bergsten, 1973, p. 3) from the standpoint of sample bias resulting from the possible homogeneity existing within and among cooperating school districts. Lord (1959) stated that

> Since schools usually differ markedly from each other in mean score, the sampling errors in the final norms table will ordinarily be large unless the number of schools in the norms sample is large; and the number of students in the norms sample

> typically has only a weak and indirect relation to the size of the sampling errors in the norms table. (p. 249)

Nevertheless, the use of cluster sampling can be viewed differently when the need is for local or special-study norms. The major purpose in constructing local/special-study norms is to provide for the best possible representation of the norm group in the sample utilized, particularly when the population to be served deviates considerably from the national "average." Therefore, the degree of homogeneity in such a group, while considered a liability in constructing national norms, becomes both the rationale for and an asset in establishing local norms.

Homogeneity with respect to spread of abilities, socio-economic levels, and size of school enrollment, can and often does exist among individual buildings in a large school district. In such cases, both school and district norms can be constructed according to need at the building level. However, when comparisons of students from various schools are needed, as in the present study, the need is clearly for pupil norms at the district level. The local norms established in the present study will be used in four large junior high schools, each of which receives students from three to four elementary schools of 500 to 900 students. Therefore, students from varying socioeconomic levels are being pooled at the seventh-grade level in each of the four secondary schools.

The sampling design utilized in the present study employed a modified two-stage sampling method. This method was designed to (a) utilize all sixth-grade students (n = 1,495, with the exception of those enrolled in classes for the Educable and Trainable Mentally Retarded); and, in so doing, to (b) achieve the greatest possible distribution of all ability levels across the separate subgroups administered Forms 1, 2, and 3 of the Gates-MacGinitie Reading Test, Survey E.

There were five major variables to consider in designing a sampling method which would achieve equivalence of norm groups for the three parallel test forms:

1. Enrollment of students among the 60 classes tested varied in pupil/teacher ratio from 22:1 to 35:1, with an average of 25:1.

2. Organizational structures for division of responsibility in teaching the various subjects included the self-contained classes, language arts blocks, and departmentalized classes. Particularly with regard to the two latter plans, the time available for testing individual classes was somewhat limited.

3. Two basic grouping methods were employed within the three organizational structures: homogeneous and heterogeneous. These groupings were based on criteria of ability levels which varied with the needs of individual building faculties or grade level committees in which the guidelines were designed.

4. A mixture of "open" and conventional plant facilities represented both environmental and attitudinal variables in relation to creating optimal classroom conditions for testing.

5. There were differing socioeconomic levels, each of which was concentrated in certain areas of the district.

The method utilized to equalize the effects of these variables across the three norm groups employed the following system whereby test forms were randomly distributed within each of the 60 classes:

1. Students were asked to "number off" orally, counting, "1, 2, 3,; 1, 2, 3, . . ." until all students in the class had arbitrarily emitted a numeral.

2. All students numbered "1" were administered all three subtests of Form 1, those numbered "2" were administered all of Form 2, and the "3s," Form 3.

This sampling method served to divide the sixth-grade population into three subgroups, each of which represented a randomly selected cross-section of a given class and each local school. In addition, each resulting norm group represented a cross-section of the entire district in relation to the effects of class size, organizational structures, basic ability-grouping methods, plant facilities ("open"/conventional), and varying socioeconomic levels. The full extent to which equivalence of academic aptitude and reading achievement levels across the three subgroups was achieved is illustrated in Tables 2 and 3.

Table 2

Means and Standard Deviations of IQ Scores Resulting from the Administration of the Short Form Test of Academic Aptitude (McGraw-Hill, 1970)

Putnam City Norm Group Number	IQ Score Mean	Standard Deviation	n =
1	105.61	12.26	493
2	105.94	13.93	484
3	105.93	13.23	456

Note. Test results from grade 6.0, Putnam City District, 1976-77 school year.

Table 3

Raw Score Means and Standard Deviations for all Forms and Subtests of Survey E of the Gates-MacGinitie Reading Test

Subtest	Form	Mean	Standard Deviation	n =
Speed and Accuracy (number correct)	1	5.888	2.106	511
	2	6.036	2.106	497
	3	5.972	2.106	468
Vocabulary	1	18.426	6.414	509
	2	20.586	6.595	495
	3	20.071	6.373	464
Comprehension	1	30.359	10.186	510
	2	29.294	10.106	496
	3	30.070	10.440	471

Note: Results of the February 1977 administration, Putnam City Schools, grade 6.5.

Instrumentation

The instrument utilized for testing and norming was the Gates-MacGinitie Reading Test, Survey E, Forms 1, 2, and 3. Each form contains three subtests: Speed and Accuracy, Vocabulary, and Comprehension.

The Gates-MacGinitie Reading Tests were nationally normed in October 1964 and April 1965 over 6,000 students in grades 7-9. These students represented 38 communities which were carefully selected on the basis of size, geographical area, average educational level, and average family income. Within each community, testing was carried out in one or more schools which were judged by the school officials to be representative of the community as a whole (Gates & MacGinitie, 1972, p. 1).

The Alternate Form and Split-Half Reliability Coefficients for each of the Survey E subtests administered in grades 7-9 are presented in Table 4. Table 5 lists the correlation coefficients between the various subtests of Survey E. Table 6 lists the correlations between the Gates-MacGinitie Reading Subtest Scores and the Lorge-Thorndike Verbal IQ for grades 7, 8, and 9. Shown in Table 7 is a comparison of Raw Score Means and Standard Deviations of the parallel test forms for grades 7-9.

The Speed and Accuracy Test consists of 36 items and is designed to be completed by the student in grades 7-9 in four minutes. Each item consists of a three- or four-sentence

Table 4

Reliability Coefficients (1964-65 Reliability Testing)

Test	Grade	Subtest	Average Raw Score Mean	Average Raw Score SD	Alternate Form Reliability	Split-Half Reliability
Survey E	7	Vocabulary	21.1	6.3	.78	.88
		Comprehension	34.9	7.0	.81	.94
		Speed and Accuracy				
		Number Attempted	13.1	4.7	.69	--
		Number Correct	11.0	2.5	.70	--
	8	Vocabulary	26.0	7.2	.80	.89
		Comprehension	33.1	9.4	.81	.93
		Speed and Accuracy				
		Number Attempted	14.3	2.9	.72	--
		Number Correct	13.1	4.3	.76	--
	9	Vocabulary	23.0	7.3	.83	.88
		Comprehension	30.1	8.8	.80	.89
		Speed and Accuracy				
		Number Attempted	16.1	3.2	.68	--
		Number Correct	15.2	3.3	.77	--

Note. From Gates-MacGinitie Reading Tests, Technical Manual No. 4017, by Gates and MacGinitie, 1972, p. 8.

Table 5

Correlations between Subtests of the Gates-MacGinitie Reading Tests

(1964-65 Reliability Testing)

Test	Grade	Vocabulary			Comprehension		Speed and Accuracy Number Attempted
		Comprehension	Speed and Accuracy		Speed and Accuracy		Speed and Accuracy Number Correct
			Number Attempted	Number Correct	Number Attempted	Number Correct	
Survey E	7	.71	.33	.61	.31	.60	.60
	8	.69	.39	.62	.27	.58	.71
	9	.73	.44	.61	.33	.55	.76

Note. From Gates-MacGinitie Reading Tests, Technical Manual No. 4017, by Gates and MacGinitie, 1972, p. 8. Copyright 1965, 1972 by Teachers College Press. Reprinted by permission.

Table 6

Correlations between Gates-MacGinitie Reading Subtest Scores and Lorge-Thorndike Verbal IQ

Test	Grade	IQ			
		Vocabulary	Comprehension	Speed and Accuracy	
				Number Attempted	Number Correct
Survey E	7	.84	.86	.37	.74
	8	.74	.76	.35	.60
	9	.84	.77	.42	.65

Note. From Gates-MacGinitie Reading Tests, Technical Manual No. 4017, by Gates and MacGinitie, 1972, p. 9.

Table 7

Comparison of Raw Score Means and Standard Deviations of Parallel Forms

(1964-65 Standardization)

Test	Grade	Subtest	Form 1			Form 2			Form 3		
			Mean	SD	N	Mean	SD	N	Mean	SD	N
E	7	Vocabulary	18.8	7.6	706	19.7	8.3	658	20.2	7.5	704
		Comprehension	30.0	12.0	706	26.8	11.3	658	29.3	11.6	701
		Speed and Accuracy Number Attempted	13.0	4.5	706	12.2	4.8	658	12.9	4.7	704
		Speed and Accuracy Number Correct	10.8	4.0	706	10.0	4.3	658	11.0	4.3	704
	8	Vocabulary	21.6	8.0	617	23.3	8.0	578	23.2	7.9	583
		Comprehension	34.3	11.4	615	32.4	10.4	577	34.0	10.1	583
		Speed and Accuracy Number Attempted	14.2	4.5	616	14.4	4.8	578	14.1	4.8	583
		Speed and Accuracy Number Correct	12.5	4.3	616	12.2	4.5	578	12.4	4.6	583
	9	Vocabulary	26.2	8.4	448	26.6	8.3	504	25.6	9.5	413
		Comprehension	39.3	10.0	448	36.5	9.6	502	35.8	11.4	410
		Speed and Accuracy Number Attempted	17.0	5.0	448	15.5	4.8	504	15.2	5.7	413
		Speed and Accuracy Number Correct	15.1	4.7	448	14.0	4.9	504	13.5	5.5	413

Note. From Gates-MacGinitie Reading Tests, Technical Manual No. 4017, by Gates and MacGinitie, 1972, p. 7.

paragraph which ends with either a question or an unfinished sentence. This paragraph is followed by four words from which the student is to choose the one which most accurately completes the meaning or answers the question. Since this test is timed, both the "number attempted" and the "number correct" are calculated, and separate norms are provided for each score.

The Vocabulary Test consists of 50 items which must be answered in 15 minutes. Each item consists of a major word for which students must find a synonym among a list of five words.

The Comprehension Test consists of 52 items which must be answered within a 25-minute time limit. Each item is written in a form which could be considered either "completion" or "cloze procedure" in that each consists of two or three statements in which a word has been deleted. For each "blank," five words are provided from which the student must choose the one which most accurately completes the meaning of the sentence. Since the deleted words in the sentences neither appear at the end of statements (as in standard completion tests) nor at regular intervals (as in the deletion of every fifth or tenth word in the original cloze procedure form), the test could be considered rather a combination of these two forms.

The answer form utilized was the Standard Answer Sheet-C (No. DS 1120-C), published by the Optical Scanning

Corporation. The use of this form aided in standardizing the scoring system to achieve greater accuracy and more dependable data.

The Short Form Test of Academic Aptitude (SFTAA; McGraw-Hill, 1970) was administered to all students in Putnam City District's grades 2, 4, 6, and 8 on October 1-3, 1976. The test-scoring and reporting of data were accomplished through the Oklahoma University Testing Services (McGraw-Hill). The individual IQ scores (total IQ) were coded on each of the OpScan scoring sheets utilized by the subjects to record answers to the Gates-MacGinitie Reading Survey Tests. These scores were read by the OpScan and punched on IBM cards along with responses to test items during the process of scoring the reading tests.

The SFTAA was standardized by administering the test in February, March, and April of 1970 to a national sample of 197,712 students in grades 1-12. It was standardized jointly with the California Achievement Tests, 1970 Edition (CAT-70). The public school standardization sample was selected by stratified random sampling with proportional allocation. The United States' population of public school students in grades 1-12 was stratified into seven geographic regions, three district sizes (based on average enrollment per grade) and four community types (urban, town, rural, and other), thus giving 84 stratification cells for the sampling design. A total of 355 public schools and 42 Catholic schools

participated in the norming process (Sullivan, Clark, & Tiegs, McGraw-Hill, 1970).

Reliability coefficients computed from the Kuder-Richardson-20 (Internal Consistency) for Level 3, sixth-grade sample, yielded the following results:

	Kuder-Richardson-20	Standard Error of Mean
Language IQ:	.91	2.81
Non-language IQ:	.90	2.50
TOTAL IQ:	.94	3.80

Pearson Product-Moment Correlations for test-retest at intervals of two weeks and 14 months for Grade 6, Level 3, yielded the following results:

	2-Week Intervals	14-Month Intervals
Language IQ:	.91	.89
Non-language IQ:	.88	.77
TOTAL IQ:	.93	.89

Selection of Test Administrators

The sixth-grade reading teachers administered both the SFTAA and the Gates-MacGinitie Reading Tests to their own reading classes in the regular classroom setting. All of these teachers had administered both tests to students two to four times over the previous four school years. This method of selection served as an aid in the effort to protect the test validity by providing the most comfortable, anxiety-free

environment for students that could be offered through approximating normal classroom conditions. Even though 86% of the administrators were certified Reading Specialists, all were given clearly stated oral and written instructions in order to provide for uniform administration of the test.

Distribution of Materials and Teacher Training

A training session was conducted in which all test administrators were given the necessary materials and instructions regarding test administration and preparation of the answer sheets (procedures for coding of the necessary information for machine-scoring and for post-test arrangement of sheets for scoring). The materials distributed included three answer forms per student; an Administrator's Test Manual; two pages of instructions regarding all of the necessary procedures; color-coded, labeled divider sheets for separating the completed answer forms for machine-scoring; and a coded answer form to use as a model (see Appendix for a copy of the Instructions for Test Administrators).

Test Administration

Administration of the tests was limited to three days in mid-February 1977. This control served four purposes:

1. The time variable was equalized in terms of grade placement and maturity level of students across the district.

2. This date coincided with the original norming and Metropolitan School Study Council test administration dates,

therefore providing a stronger basis on which to compare norms.

3. Pre-enrollment begins in Putnam City Junior High Schools on April 1st each year. It is at this time that all currently enrolled students are placed in appropriate classes. The mid-February testing provided the time allowance needed to score tests, compute norms, and return the results to the junior high school teachers and counselors with the proper instructions regarding interpretation and use of the data.

4. Many of the sixth-grade classes were departmentalized. Under such conditions, teachers were limited to 45 to 75 minutes of actual teaching/testing time per day with each class, depending on the number of teachers per building and the number of subjects taught by each. With a total test time of 44 minutes, and additional time needed between subtests for directions, a three-day limit allowed ample time for administering the tests to all classes.

Tests which could not be administered because of student absences had to be made up during the three-day limit or missed entirely. This accounts for the variance in within-sample sizes for each of the groups represented by Forms 1, 2, and 3, presented in Table 8.

Table 8

Sample Size for Each Subtest of the

Gates-MacGinitie Reading Tests

Survey E, Forms 1, 2, and 3

	Speed and Accuracy	Vocabulary	Comprehension
Form 1	511	509	510
Form 2	497	495	496
Form 3	468	464	471

Note. Samples drawn from the 1,495 eligible sixth-grade subjects, Putnam City Schools, 1977.

Preparation for Scoring and Analysis

The IQ scores obtained from the October 1976 administration of the Short Form Test of Academic Aptitude (McGraw-Hill, 1970 ed.) were machine-coded and recorded on the 4,421 OpScan Answer Sheets after the test was administered. The use of this procedure facilitated computer analysis of the data.

Scoring and Recording of the Data

All answer forms were scored on the Optical Scanning Machine. During the scanning process, data cards were being punched with students' names, IQ scores, and responses to each test item. This process yielded one card per student for the Speed and Accuracy Test, and two cards per student

for each of the Vocabulary and Comprehension Tests. Actual test scoring was achieved through the use of the computer.

Treatment of the Data

The major portion of the data treatment was completed through use of the computer at the University of Oklahoma, Norman, Oklahoma. A basic test analysis program (J803) was utilized to calculate the following statistics: mean, standard deviation, percentile rank, class rank, raw score, z-score, T-score, item analysis, frequency distribution, standard error of measurement, and the Kuder-Richardson Formula 20 (reliability coefficient). Separate programming yielded raw scores, means, and standard deviations for "number attempted" scores for the Speed and Accuracy subtests.

Additional manual calculations yielded percentile bands and stanine scales for all raw score data, as well as percentile ranks, z-scores, and T-scores for Speed and Accuracy, "number attempted." An Item Factor Analysis was conducted using the Comprehension subtest items for Forms 1, 2, and 3. The computer program utilized to analyze these item data (Biomedical Factor Analysis Computer Program No. BMDO8M, Dixon, 1975) yielded means and standard deviations of the IQ score data for each of the three norm groups. The results of these analyses are presented in Table 2.

Several formulae were utilized in an attempt to establish a non-spurious lower-bounds estimate of reliability for each

of the subtests. The Kuder-Richardson Formula 20 (K-R 20) was initially applied to all subtest data through use of the computer program J803. The resulting reliability coefficients are presented in Table 9.

Although test publishers commonly use the K-R 20 to obtain single-trial estimates of reliability, these reliability coefficients are spuriously inflated when applied to data obtained from the administration of highly speeded tests (Cronbach & Warrington, 1951, p. 168). In an effort to compensate for the effects of the degree of speededness and resulting interdependence of test items on the reliability coefficient, Gulliksen (1950, pp. 259-269) derived three lower-bounds formulae to obtain single-trial estimates of reliability for speeded tests. Cronbach and Warrington (1951, pp. 176-177) refined and adjusted these formulae to develop two additional formulae in which the greatest and least number of items completed by any given subject were taken into consideration.

Guilford (1954) stated that the split-half reliability formulae based on an odd-even division of test items "overestimates reliability when there is an appreciable amount of speeding . . . and should thus be regarded as an upper-bound estimate under this condition" (p. 391). As an alternate solution, Guilford (1954, p. 392) recommended the use of Gulliksen's error score formula in which the ratio of mean number of items attempted at the end of the test (M_u) to the

Table 9

Reliability Coefficients for the

1977 Reliability Testing

Gates-MacGinitie Reading Test,

Survey E, Forms 1, 2, and 3

Subtest	Form	K-R 20 Reliability Coefficient
Speed and Accuracy Number Correct	1	.679[a]
	2	.683[a]
	3	.699[a]
Vocabulary	1	.797
	2	.813
	3	.795
Comprehension	1	.911
	2	.909
	3	.918

[a]Due to the effects of speededness, this reliability coefficient may be spuriously inflated.

Note. Putnam City Schools, grade 6.5.

variance of the total error score (expressed as σ_x^2, the variance of all wrong responses plus items omitted) is subtracted from the odd-even reliability coefficient (stepped up by the Spearman-Brown formula). Two criteria have been established to determine the appropriateness of the resulting calculations: (a) "when the ratio of M_u to σ_x^2 is greater than .2 to .3, the test is so speeded that even this formula should not be used" (Guilford, 1954, p. 392); and (b) when the standard deviation of u (unattempted items) is much greater than .2 to .3 of the standard deviation of w (items answered wrongly), "a split-half correlation is an unsafe basis for estimating the test reliability" (Gulliksen, 1950, p. 235).

When the Gulliksen (1950, pp. 259-269) and Cronbach and Warrington (1951, pp. 176-177) formulae were applied to the Speed and Accuracy subtest data for the Putnam City norm samples, the results exceeded the limits normally placed on reliability coefficients (i.e., between +1 and -1), thus yielding spuriously high reliability coefficients. The reliabilities obtained through use of the Gulliksen error score formula showed that neither the Guilford nor the Gulliksen criteria could be met with the Speed and Accuracy test data. Evidence of contributing factors was found in the item analyses and calculations of "number attempted" scores from the Speed and Accuracy subtests which revealed that, of the 36 test items comprising each of the parallel

forms, the percentages of total test items (a) answered correctly, (b) answered wrongly, and (c) unattempted were 34%, 44%, and 22%, respectively, for Form 1; 25%, 44%, and 31%, respectively, for Form 2; and 25%, 50%, and 25%, respectively, for Form 3.

Testing the Hypotheses

Hypotheses 1-4 were tested with the *F*-test for the Analysis of Variance testing statistic (ANOVA [Gellman, 1973, pp. 188-91]). Hypotheses 5-7 were tested with the *z*-test for an hypothesis about a mean (two-sample test for large samples with unequal variances [Gellman, 1973, p. 197]). Hypotheses 8-10 were tested with the one-sample *z*-test for an hypothesis about a mean (Gellman, 1973, p. 197). The results of testing the hypotheses are discussed in Chapter IV. The conclusions drawn from these findings and recommendations for further study are discussed in Chapter V.

CHAPTER IV

RESULTS OF THE DATA ANALYSIS

Introduction

The problem researched in this study was to establish local norms for Putnam City sixth-grade students on the Gates-MacGinitie Reading Test (GMRT), Survey E, Forms 1, 2, and 3. Each of the test forms contains three subtests: Speed and Accuracy, Vocabulary, and Comprehension.

A modified cluster sampling procedure was utilized in which all sixth-grade students were tested. The three test forms were randomly distributed through a "counting-off" system utilized in each classroom unit. The three samples resulting from this random distribution of test forms were treated as separate norm groups. Throughout this paper, the samples were designated by the test form (1, 2, or 3) administered in each case.

In order to estimate the equivalence of academic aptitude among the three samples, results of the 1976-77 administration of the Short Form of Academic Aptitude (SFTAA, McGraw-Hill, 1970 ed.) were coded on each student's answer sheets for Speed and Accuracy, Vocabulary, and Comprehension.

During the computer data processing phase of this study, means and standard deviations of total IQ scores were calculated for each sample, thus providing the necessary statistics for analysis.

Results of Testing the Hypotheses

Ten hypotheses were formulated and tested in order to determine and/or establish (a) the need for specialized local norms for this particular sixth-grade population; (b) the equivalence of academic aptitude among the three samples tested; and (c) the equivalence of the three test forms in measuring reading achievement. The results of testing each hypothesis are discussed in the remaining pages of this section.

The first four hypotheses were tested at the .05 level of significance with the F-test for the Analysis of Variance testing statistic (ANOVA; Gellman, 1973, pp. 188-91). The results of the statistical calculations for the ANOVA are presented in Table 10.

Ho_1: The results of testing the first hypothesis indicated that no statistically significant differences existed among the total IQ score means (SFTAA, McGraw-Hill, 1970 ed.) for the Putnam City norm samples formed by the random distribution and administration of the Gates-MacGinitie Reading Test, Survey E, Forms 1, 2, and 3, F (2/1430) = .00020373, $p < .05$. These results would not allow the first null hypothesis to be rejected.

Table 10

Results of Testing the First Four Hypotheses with the F-Test for the Analysis of Variance Testing Statistic

	Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Value
Ho_1	Between-Groups	.0706	2	.0353	.00020373*
	Within-Groups	247773.9513	1430	173.2684974	
*Not significant at the .05 level.					
Ho_2	Between-Groups	.01102025	2	.005510125	.0012398336*
	Within-Groups	6546.373596	1473	4.444245483	
*Not significant at the .05 level.					

Table 10 (Continued)

	Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Value
Ho_3	Between-Groups	2.546823075	2	1.273411538	.0304256788*
	Within-Groups	61314.9148	1465	41.85318416	
*Not significant at the .05 level.					
Ho_4	Between Groups	.6066476926	2	.3033238463	.0029750372*
	Within-Groups	150283.6157	1474	101.95632	
*Not significant at the .05 level.					

Ho_2: The results of testing the second hypothesis indicated that no statistically significant differences existed among the Speed and Accuracy subtest raw score means (number correct) for the three Putnam City norm samples represented by Forms 1, 2, and 3 of the Gates-MacGinitie Reading Test, Survey E, F (2/1473) = .0012398336, $p < .05$. These results would not allow the second null hypothesis to be rejected.

Ho_3: The results of testing the third hypothesis indicated that no statistically significant differences existed among the Vocabulary subtest raw score means for the Putnam City norm samples represented by Forms 1, 2, and 3 of the Gates-MacGinitie Reading Test, Survey E, F (2/1465) = .0304256788, $p < .05$. These results would not allow the third null hypothesis to be rejected.

Ho_4: The results of testing the fourth null hypothesis indicated that no statistically significant differences existed among the Comprehension subtest raw score means for the Putnam City norm samples represented by Forms 1, 2, and 3 of the Gates MacGinitie Reading Test, Survey E, F (2/1474) = .0029750372, $p < .05$. These results would not allow the fourth hypothesis to be rejected.

Hypotheses 5-7 were tested at the .05 level of significance with the z-test for an hypothesis about a mean (two-sample test for large samples with unequal variances [Gellman, 1973, p. 197]). The statistical test for each hypothesis employed a procedure whereby this z-test was

applied to means of matched reading test form data for the two norm groups being compared (i.e., the mean of Form 1 data for the Putnam City sample was compared with the mean of Form 1 data for the seventh-grade portion of the original norm group; Form 2 was compared with Form 2; and Form 3 with Form 3 for each of the two norm samples).

Ho_5: The results of testing the fifth hypothesis indicated that statistically significant differences were found in comparing the Putnam City raw score means for Speed and Accuracy (number correct) with those of the seventh-grade portion of the original norm group. The seventh-grade mean for each of the test forms was found to be significantly greater than the local raw score means for identical tests. The calculated values of z were found as follows: $z = 27.74547411$, $p > + 1.96$; $z = 20.60108528$, $p > + 1.96$; and $z = 26.59565415$, $p > + 1.96$ for Forms 1, 2, and 3, respectively. Ho_5 was therefore rejected at the .05 level of significance.

Ho_6: The results of testing the sixth hypothesis indicated that no statistically significant differences were found in the Vocabulary raw score means on Forms 1 and 3 when comparing the Putnam City test results with those of the seventh-grade portion of the original norm group. The calculated values of z were found as follows: $z = .92738899$, $p < + 1.96$ for Form 1; and $z = .3152600611$, $p < + 1.96$ for Form 3. However, a statistically significant difference was

found when comparing Vocabulary raw score means of Form 2 for the respective groups; calculated value of z = -2.01905228, p < -1.96, thus indicating that the raw score mean for the Putnam City sample was significantly greater than that of the seventh-grade students in the original norm group. Ho_6 was therefore rejected at the .05 level.

Ho_7: The results of testing the seventh hypothesis indicated that no statistically significant differences were found in the Comprehension raw score means on Forms 1 and 3 when comparing the Putnam City test results with those of the seventh-grade portion of the original norm group. The calculated values of z were found as follows: z = -0.5624456087, p > -1.96 for Form 1; and z = -1.18340701, p > -1.96 for Form 3. However, a statistically significant difference was found when comparing Comprehension raw score means of Form 2 for the respective groups; calculated value of z = -3.943520447, p < -1.96, thus indicating that the raw score mean for the Putnam City sample was significantly greater than that of the seventh-grade students in the original norm group. Ho_7 was therefore rejected at the .05 level.

Hypotheses 8-10 were tested at the .05 level of significance with the one-sample z-test for an hypothesis about a mean (Gellman, 1973, p. 197). The statistical test for each hypothesis employed a procedure whereby this z-test was utilized to compare each Putnam City raw score mean for Form 1, 2, or 3 of a given subtest with the interpolated raw score

equivalent of the grade equivalent score 6.5 from the original norm scale. The Putnam City norm groups represented by the random distribution of the Gates-MacGinitie Reading Test, Survey E, Forms 1, 2, and 3 were administered the respective test forms at grade placement 6.5. Their respective raw score means for each subtest and test form would therefore be considered equivalent to a grade equivalent score of 6.5.

Ho_8: The results of testing the eighth hypothesis indicated that statistically significant differences were found in the Putnam City raw score means for Speed and Accuracy (number correct) and the interpolated raw score equivalents of the 6.5 grade equivalent scores for all test forms, based on the original norm data. The calculated values for z were found as follows: $z = -41.45380045$, $p < -1.96$; $z = -39.31531205$, $p < -1.96$; and $z = -38.80846926$, $p < -1.96$ for Forms 1, 2, and 3, respectively. In each comparison made, the interpolated raw score equivalent of 6.5 on the original norm scale was found to be significantly greater than the mean for Putnam City students tested at grade placement 6.5. Ho_8 was therefore rejected at the .05 level.

Ho_9: The results of testing the ninth hypothesis indicated that statistically significant differences were found in the Putnam City raw score means for the Vocabulary subtest and the interpolated raw score equivalents of the

6.5 grade equivalent scores for all test forms, based on the original norm data. The calculated values for z were found as follows: $z = 2.201933855$, $p > + 1.96$; $z = 9.39872433$, $p > + 1.96$; and $z = 7.675951219$, $p > + 1.96$ for Forms 1, 2, and 3, respectively. In each of the comparisons made, the raw score means for Putnam City students tested at grade placement 6.5 was found to be significantly greater than the interpolated raw score equivalent of 6.5 on the original norm scale. Ho_9 was therefore rejected at the .05 level.

Ho_{10}: The results of testing the tenth hypothesis indicated that statistically significant differences were found in the Putnam City raw score means for the Comprehension subtest and the interpolated raw score equivalents of the 6.5 grade equivalent scores for all test forms, based on the original norm data. The calculated values for z were found as follows: $z = 7.447172611$, $p > + 1.96$; $z = 5.055393409$, $p > + 1.96$; and $z = 6.38187554$, $p > + 1.96$ for Forms 1, 2, and 3, respectively. In each of the comparisons made, the raw score means for Putnam City students tested at grade placement 6.5 was found to be significantly greater than the interpolated raw score equivalent of 6.5 on the original norm scale. Ho_{10} was therefore rejected at the .05 level.

Addressing the Related Questions

The problem researched in this study was to establish local/special-study norms for the Putnam City sixth-grade

population on the Gates-MacGinitie Reading Test, Survey E, Forms 1, 2, and 3. This test was constructed and nationally normed for students in grades 7-9 (Gates and MacGinitie, 1964). In 1965, Gates and MacGinitie established special-study norms on this test for students in the 70 school districts comprising the Metropolitan School Study Council (MSSC). These students' academic aptitude and achievement levels exceeded those of the national norm group to the extent that the original norms were considered inappropriate for use in the MSSC schools.

The need for special-study norms for the Putnam City students was based on evidence of the consistently superior performance levels achieved by these students when compared with those achieved by national norm groups (e.g., the California Test of Basic Skills data presented in Table 1). Two related questions were explored in order (a) to determine the degree to which the existing norms constructed for the Gates-MacGinitie Reading Test, Survey E, were appropriate for use with Putnam City sixth-grade students; and (b) to establish statistical evidence of the need for local norms. Each of the questions was considered separately in regard to the resulting Speed and Accuracy, Vocabulary, and Comprehension subtest data:

1. Were there statistically significant differences in the raw score means obtained for the Putnam City samples tested at grade placement 6.5 and the raw score equivalents

of the interpolated grade equivalent score of 6.5 on the original norm scales?

2. Would a comparison of the Putnam City and Metropolitan School Study Council norms established for Survey E at grade 6.5 (Gates & MacGinitie, 1965) confirm the relevance of MSSC norms for Putnam City use?

Hypotheses 8-10 were stated and tested in order to establish the evidence needed to answer the first question. The relevant findings were presented in the preceding section.

Exploration of the second question necessitated a comparison of the Putnam City subtest data with that obtained from the MSSC norm group tested at grade placement 6.5. The derived scores presented in the MSSC norm study were expressed in terms of normalized standard scores and percentiles representing the weighted means for the combined forms of each subtest. The raw score equivalents of 50 on the standard score distributions for MSSC students tested at grade placement 6.5 were utilized for comparison.

The weighted means for the combined forms of each subtest were calculated from the Putnam City test data. Comparisons of the relevant norm group data were tested at the .05 level with the one-sample z-test for an hypothesis about a mean (Gellman, 1973, p. 197). The following results reveal the statistically significant differences found in each of the comparisons: (a) Speed and Accuracy (number attempted), calculated $z = -41.2010414$, $p < -1.96$; (b) Speed and Accuracy

(number correct), calculated $z = -183.0727995$, $p < -1.96$; (c) Vocabulary, calculated $z = -19.0440209$, $p < -1.96$; and (d) Comprehension, calculated $z = -17.2295255$, $p < -1.96$.

These findings revealed that the MSSC norm groups' total weighted means for the combined forms of each subtest were statistically significantly greater than the weighted means of comparable subtest data achieved by the Putnam City norm groups.

Construction of the Norm Scales

Introduction

The purpose of establishing norms is to provide standard and other derived score scales with which to interpret raw scores. Construction of local and/or special study norms is accomplished for the purpose of bringing the greatest possible relevance to interpreting the raw scores of a specified group on a given measure of achievement.

The local norms established in the present study provide for adequate interpretation of the Gates-MacGinitie Reading Test, Survey E, Forms 1, 2, and 3 for Putnam City District's sixth-grade students. These norms are expressed in percentile ranks, percentile bands, z-scores, T-scores, and stanines. Grade equivalent scales, normally included in national norming procedures, were purposely omitted. The reason for this omission is that all students tested were at grade placement 6.5 at the time. Extrapolation of grade equivalent scores

from a single point would lead to data so arbitrary as to be errant and/or invalid.

The remaining pages of this chapter contain (a) formulae for and/or explanations of the statistical procedures utilized in obtaining the relevant norm scales; and (b) the norms tables constructed for use in Putnam City School District. In order to facilitate the use of these norms, scales representing different types of derived scores for the same set of data are often combined in one table. Furthermore, to provide for clarity of presentation and ease of norm scale comparisons among the three test forms, the statistical procedures and norms tables are presented in separate sections.

Explanation of the Statistical Procedures

The formulae utilized manually and by computer to calculate the statistics necessary for constructing the norm scales are presented as follows:

1. Percentile Ranks

   a. Grouped Data (Gellman, 1973, p. 87)

$$\text{Percentile Rank} = \frac{cf + \frac{X-LL}{i}(f)}{N} \times 100$$

where: cf = the cumulative frequency of the interval below the one that includes the score in question

X = the score

LL = the lower exact limit of the interval that includes the score

i = the interval size

f = the frequency of the interval that includes the score

N = the total number of students

b. Ungrouped Data (derived from the formula for grouped data)

$$\text{Percentile Rank} = \frac{cf_x - \frac{f_x}{2}}{N}$$

where: cf_x = cumulative frequency at the point of the score in question

f_x = frequency of occurrence of the score in question

N = total number of scores in the distribution

2. The Kuder-Richardson Formula 20 (Guilford, 1954, p. 380) used to obtain an estimate of reliability for the Vocabulary and Comprehension subtests:

$$r_{tt} = \left(\frac{n}{n-1}\right)\left(\frac{\sigma_t^2 - \Sigma pq}{\sigma_t^2}\right)$$

where: r_{tt} = the estimated reliability of a test, based on item statistics

n = number of items in the test

p = proportion of correct responses to each item in turn (or proportion of examinees responding in the keyed manner)

q = proportion of people who answered an item incorrectly ($q = 1 - p$)

pq = variance of a single item scored dichotomously (right or wrong)

Σ = summation sign indicating that pq is summed over all items

σ_t^2 = variance of the total test

3. Standard Error of Measurement (Mehrens & Lehmann, 1973, p. 107):

$$S_e = S_x \sqrt{1 - r_{xx}}$$

where: S_e = Standard Error of Measurement

S_x = Raw Score Standard Deviation

r_{xx} = correlation coefficient (K-R 20 or Pearson Product Moment Correlation can be used)

4. Percentile Band:

By definition, the percentile band for any given score is a range of percentile ranks, the upper and lower limits of which are equivalent to obtained scores which are one standard error of measurement above and below the score in question. Procedural steps for constructing percentile bands for a given score include the following:

a. Subtract one standard error of measurement from the score in question.

b. Find the percentile rank which is equivalent to this score, and record it as the lower limit of the percentile band.

c. Starting again with the original score for which a percentile band is to be constructed, add one standard error of measurement unit.

d. Find the equivalent percentile rank of the score obtained in Step c. Record this value as the upper limit of the percentile band.

5. $\underline{z}$-scores were calculated from the following formula:

$$\underline{z} = \frac{X - \bar{X}}{SD}$$

where: $\underline{z}$ = the "simplest form of standard score . . . an expression of the deviation of a score from the mean score of the group in relation to the standard deviation of the scores of the group" (Mitchell, 1976, p. 7). The mean of the $\underline{z}$ score is zero; the standard deviation is 1.

X = the score in question

$\bar{X}$ = the mean of the distribution

SD = the standard deviation of the distribution

6. T-scores were calculated from the following formula (Thorndike & Hagen, 1969, p. 227):

$$T = 10\left[\frac{X - \bar{X}}{SD}\right] + 50 \quad \text{or} \quad T = 10\,(\underline{z}) + 50$$

where: T = a standard score unit having a mean of 50 and a standard deviation of 10

X = the specific raw score in question

$\bar{X}$ = the mean of the distribution

SD = the standard deviation of the distribution

$\underline{z}$ = the standard $\underline{z}$-score, obtained with the formula: $\underline{z} = \frac{X - \bar{X}}{SD}$

7. Stanine (or standard nine) scales have values from 1 to 9. The mean of the stanine is 5, and the standard deviation is 2 (Mitchell, 1976, p. 8).

Each stanine is 1/2 standard deviation in width. The middle of stanine 5 is equivalent to the mean of

a given distribution. The boundaries for each point on the stanine scale, listed below, are invariable, regardless of the manner in which scores are distributed.

Stanine	Boundary of point on the distribution:
1	From below -1 3/4 σ to any point below -1 3/4 σ
2	From -1 3/4 σ to -1 1/4 σ
3	From -1 1/4 σ to -3/4 σ
4	From -3/4 σ to -1/4 σ
5	From -1/4 σ to +1/4 σ
6	From +1/4 σ to +3/4 σ
7	From +3/4 σ to +1 1/4 σ
8	From +1 1/4 σ to +1 3/4 σ
9	From +1 3/4 σ to any point above +1 3/4 σ

The Norms Tables

The norms tables contained in this section are organized as follows:

1. Tables 11-22 exhibit the following scales: Obtained Raw Score, Frequency, Cumulative Frequency, and Percentile Rank. Each table exhibits relevant data from one form of one subtest. Three tables each present norm scales for Speed and Accuracy (number attempted and number correct), Vocabulary, and Comprehension.

2. Each of Tables 23-26 presents a comparison of Obtained Raw Scores with their accompanying Percentile Bands for all three forms of a given subtest. These comparisons are presented separately for Speed and Accuracy (number attempted and number correct), Vocabulary, and Comprehension.

3. Tables 27-38 exhibit the following scales: Raw Score, z-Score, T-Score, and Stanine. Presented in each table are relevant data from one form of one subtest. Three tables each exhibit norm scales for Speed and Accuracy (number attempted and number correct), Vocabulary, and Comprehension.

The norm scales presented in Tables 27 through 38 include calculated z- and T-score values for all obtainable raw scores. Interpolated and/or extrapolated raw score values inserted in these scales are enclosed in parentheses to indicate that these scores were not among the obtained score data collected for the present study. The purpose for expanding the local norm scales to include the full range of obtainable raw and derived scores was to facilitate the use of these norm data at the classroom, building, and district levels.

Table 11

Frequency Distributions and Percentile Ranks for the Gates-MacGinitie Reading Test, Survey E, Form 1 Speed and Accuracy, Number Attempted

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
36	3	511	99+
28	1	508	99
26	1	507	99
25	1	506	99
24	2	505	99
23	2	503	98
22	1	501	98
21	4	500	98
20	6	496	97
19	8	490	95
18	18	482	93
17	15	464	89
16	37	449	84
15	34	412	77
14	51	378	69
13	39	327	60
12	44	288	52
11	64	244	42
10	53	180	30
9	49	127	20
8	35	78	12
7	17	43	7
6	13	26	4
5	5	13	2
4	6	8	1
3	1	2	--
1	1	1	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 12

Frequency Distributions and Percentile Ranks for the Gates-MacGinitie Reading Test, Survey E, Form 2

Speed and Accuracy, Number Attempted

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
36	1	497	99+
34	1	496	99+
30	1	495	99+
25	2	494	99
24	1	492	99
23	2	491	99
22	1	489	98
21	5	488	98
20	9	483	96
19	5	474	95
18	23	469	92
17	24	446	87
16	25	422	82
15	43	397	76
14	30	354	68
13	41	324	61
12	60	283	51
11	55	223	39
10	68	168	27
9	36	100	17
8	22	64	11
7	18	42	7
6	9	24	4
5	9	15	2
4	5	6	1
3	1	1	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 13

Frequency Distributions and Percentile Ranks for the Gates-MacGinitie Reading Test, Survey E, Form 3 Speed and Accuracy, Number Attempted

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
36	1	468	99+
33	1	467	99+
28	1	466	99+
27	1	465	99
25	1	464	99
24	2	463	99
23	4	461	98
22	3	457	97
21	1	454	97
20	7	453	96
19	8	446	94
18	14	438	92
17	17	424	89
16	25	407	84
15	39	382	78
14	38	343	69
13	45	305	60
12	62	260	49
11	52	198	37
10	45	146	26
9	40	101	17
8	32	61	10
7	10	29	5
6	9	19	3
5	8	10	1
4	2	2	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 14

Frequency Distributions and Percentile Ranks for the Gates-MacGinitie Reading Test, Survey E, Form 1

Speed and Accuracy, Number Correct

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
13	3	511	99+
12	2	508	99
11	3	506	99
10	12	503	97
9	26	491	94
8	67	465	84
7	79	398	70
6	99	319	53
5	99	220	33
4	56	121	18
3	40	65	9
2	15	25	3
1	7	10	1
0	3	3	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 15

Frequency Distributions and Percentile Ranks for the Gates-MacGinitie Reading Test, Survey E, Form 2

Speed and Accuracy, Number Correct

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
13	2	497	99+
12	1	495	99+
11	8	494	99
10	13	486	97
9	40	473	91
8	55	433	82
7	74	378	69
6	91	304	52
5	111	213	32
4	48	102	16
3	32	54	8
2	16	22	3
1	5	6	1
0	1	1	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 16

Frequency Distributions and Percentile Ranks for the Gates-MacGinitie Reading Test, Survey E, Form 3

Speed and Accuracy, Number Correct

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
16	1	468	99+
15	1	467	99+
14	1	466	99+
13	1	465	99
12	2	464	99
11	5	462	98
10	11	457	97
9	26	446	93
8	51	420	84
7	69	369	72
6	98	300	54
5	94	202	33
4	60	108	17
3	31	48	7
2	15	17	2
1	2	2	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 17

Frequency Distributions and Percentile Ranks for the Gates-MacGinitie Reading Test, Survey E, Form 1

Vocabulary Subtest

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
39	1	509	99+
38	1	508	99+
37	1	507	99+
35	1	506	99
34	2	505	99
33	3	503	99
32	6	500	98
31	10	494	96
30	1	484	95
29	2	483	95
28	7	481	94
27	14	474	92
26	16	460	89
25	16	444	86
24	22	428	82
23	25	406	77
22	26	381	72
21	32	355	67
20	27	323	61
19	23	296	56

Table 17 (Continued)

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
18	46	273	49
17	37	227	41
16	33	190	34
15	32	157	28
14	29	125	22
13	12	96	18
12	13	84	15
11	19	71	12
10	6	52	10
9	14	46	8
8	9	32	5
7	9	23	4
6	3	14	3
5	1	11	2
4	3	10	2
3	1	7	1
2	3	6	1
1	2	3	--
0	1	1	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 18

Frequency Distributions and Percentile Ranks for the Gates-MacGinitie Reading Test, Survey E, Form 2

Vocabulary Subtest

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
43	1	495	99+
42	1	494	99+
40	1	493	99+
38	3	492	99
37	1	489	99
36	1	488	99
35	4	487	98
34	4	483	97
33	5	479	96
32	13	474	94
31	8	461	92
30	6	453	91
29	12	447	89
28	11	435	87
27	18	424	84
26	21	406	80
25	17	385	76
24	30	368	71
23	16	338	67
22	26	322	62

Table 18 (Continued)

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
21	21	296	58
20	41	275	51
19	35	234	44
18	38	199	36
17	37	161	29
16	30	124	22
15	17	94	17
14	16	77	14
13	10	61	11
12	18	51	9
11	5	33	6
10	7	28	5
9	9	21	3
8	5	12	2
6	3	7	1
5	1	4	1
4	1	3	1
3	1	2	--
0	1	1	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 19

Frequency Distributions and Percentile Ranks for the Gates-MacGinitie Reading Test, Survey E, Form 3

Vocabulary Subtest

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
37	1	464	99+
35	1	463	99+
34	6	462	99+
33	3	456	98
32	8	453	97
31	6	445	95
30	11	439	93
29	13	428	91
28	12	415	88
27	10	403	86
26	15	393	83
25	23	378	79
24	23	355	74
23	27	332	69
22	35	305	62
21	31	270	55
20	31	239	48
19	18	208	43
18	31	190	38
17	36	159	30

Table 19 (Continued)

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
16	19	123	25
15	15	104	21
14	19	89	17
13	13	70	14
12	14	57	11
11	12	43	8
10	8	31	6
9	5	23	4
8	3	18	4
7	3	15	3
6	5	12	2
5	4	7	1
4	1	3	1
2	2	2	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 20

Frequency Distributions and Percentile Ranks for the Gates-MacGinitie Reading Test, Survey E, Form 1

Comprehension Subtest

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
51	3	510	99+
50	3	507	99
49	4	504	98
48	3	500	98
47	4	497	97
46	12	493	96
45	14	481	93
44	8	467	91
43	11	459	89
42	13	448	87
41	17	435	84
40	13	418	81
39	21	405	77
38	12	384	74
37	17	372	71
36	20	355	68
35	20	335	64
34	15	315	60
33	16	300	57
32	21	284	54
31	19	263	50
30	17	244	46
29	15	227	43
28	17	212	40
27	17	195	37
26	12	178	34

Table 20 (Continued)

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
25	16	166	31
24	15	150	28
23	13	135	25
22	16	122	22
21	13	106	20
20	14	93	17
19	8	79	15
18	10	71	13
17	7	61	11
16	6	54	10
15	7	48	9
14	12	41	7
13	5	29	5
12	3	24	4
11	5	21	4
10	8	16	2
9	1	8	2
7	1	7	1
6	4	6	1
5	1	2	--
2	1	1	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 21

Frequency Distributions and Percentile Ranks for the Gates-MacGinitie Reading Test, Survey E, Form 2. Comprehension Subtest

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
51	3	496	99+
50	1	493	99
49	1	492	99
48	5	491	99
47	8	486	97
46	6	478	96
45	5	472	95
44	11	467	93
43	17	456	90
42	12	439	87
41	13	427	85
40	22	414	81
39	11	392	78
38	15	381	75
37	4	366	73
36	8	362	72
35	22	354	69
34	16	332	65
33	15	316	62
32	14	301	59
31	11	287	57
30	13	276	54
29	17	263	51
28	27	246	47
27	16	219	43
26	13	203	40

Table 21 (Continued)

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
25	19	190	36
24	22	171	32
23	14	149	29
22	11	135	26
21	20	124	23
20	10	104	20
19	15	94	17
18	15	79	14
17	10	64	12
16	9	54	10
15	9	45	8
14	4	36	7
13	5	32	6
12	8	27	5
11	5	19	3
10	6	14	2
9	4	8	1
8	1	4	1
7	2	3	--
6	1	1	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 22

Frequency Distributions and Percentile Ranks for the Gates-MacGinitie Reading Test, Survey E, Form 3

Comprehension Subtest

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
51	1	471	99+
49	1	470	99+
48	5	469	99
47	8	464	98
46	9	456	96
45	8	447	94
44	13	439	92
43	10	426	89
42	12	416	87
41	15	404	84
40	20	389	81
39	11	369	77
38	18	358	74
37	14	340	71
36	18	326	67
35	22	308	63
34	11	286	60
33	17	275	57
32	21	258	53
31	18	237	48
30	13	219	45
29	13	206	42
28	18	193	39
27	18	175	35
26	11	157	32
25	2	146	31

Table 22 (Continued)

Obtained Raw Score	Frequency	Cumulative Frequency	Percentile Rank
24	6	144	30
23	14	138	28
22	9	124	25
21	12	115	23
20	11	103	21
19	11	92	18
18	13	81	16
17	9	68	14
16	7	59	12
15	8	52	10
14	12	44	8
13	6	32	6
12	5	26	5
11	4	21	4
10	6	17	3
9	1	11	2
8	5	10	2
7	1	5	1
6	2	4	1
3	1	2	--
0	1	1	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 23

Raw Scores and Corresponding Percentile Bands for the Gates-MacGinitie Reading Test, Survey E, Forms 1-3

Speed and Accuracy, Number Attempted

Obtained Raw Score	Percentile Bands		
	Form 1	Form 2	Form 3
36	99-99+	99+	99+
34	99-99+	99+	99+
33	99-99+	99+	99+
30	99-99+	99-99+	99+
28	99	99-99+	99-99+
27	99	99-99+	99-99+
26	99	99-99+	99
25	98-99	99-99+	98-99
24	98-99	98-99	97-99
23	97-99	98-99	97-99
22	95-99	96-99	96-99
21	93-99	95-99	94-98
20	89-98	92-98	92-97
19	84-98	87-98	89-97
18	77-98	82-96	84-96
17	69-97	76-95	78-94
16	60-95	68-92	69-92
15	52-93	61-87	60-89
14	42-89	51-82	49-84
13	30-84	39-76	37-78

Table 23 (Continued)

Obtained Raw Score	Percentile Bands		
	Form 1	Form 2	Form 3
12	20-77	27-68	26-69
11	12-69	17-61	17-60
10	7-60	11-51	10-49
9	4-52	7-39	5-37
8	2-42	4-27	3-26
7	.1-30	2-17	1-17
6	0-20	1-11	0-10
5	0-12	0- 7	0- 5
4	0- 7	0- 4	0- 3
3	0- 4	0- 2	0- 1
2	0- 2	0- 1	--
1	0- 1	--	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 24

Raw Scores and Corresponding Percentile Bands for the Gates-MacGinitie Reading Test, Survey E, Forms 1-3

Speed and Accuracy, Number Correct

Obtained Raw Score	Percentile Bands		
	Form 1	Form 2	Form 3
16	99+	99+	99+
15	99+	99+	99+
14	99+	99+	99
13	99	99	99
12	98-99	99	98-99
11	97-99	97-99	97-98
10	94-99	91-99	93-98
9	84-97	82-97	84-97
8	70-94	69-91	72-93
7	53-84	52-82	54-84
6	33-70	32-69	33-72
5	18-53	16-52	17-54
4	9-33	8-32	7-33
3	3-18	3-16	2-17
2	1- 9	0- 8	0- 7
1	0- 3	0- 3	0- 2

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 25

Raw Scores and Corresponding Percentile Bands for the Gates-MacGinitie Reading Test, Survey E Vocabulary, Forms 1, 2, and 3

Obtained Raw Score	Percentile Bands		
	Form 1	Form 2	Form 3
43	99+	99+	99+
42	99+	99-99+	99+
41	99+	99-99+	99+
40	99+	99-99+	99+
39	99-99+	99-99+	99+
38	99-99+	98-99+	99+
37	99-99+	97-99+	99-99+
36	99-99+	96-99	98-99+
35	98-99+	94-99	97-99+
34	96-99+	92-99	95-99+
33	95-99	91-99	93-99+
32	95-99	89-98	91-99+
31	94-99	87-97	88-99
30	92-99	84-96	86-98
29	89-98	80-94	83-97
28	86-96	76-92	79-95
27	82-95	71-91	74-93
26	77-95	67-89	69-91
25	72-94	62-87	62-88
24	67-92	58-84	55-86
23	61-89	51-80	48-83

Table 25 (Continued)

Obtained Raw Score	Percentile Bands		
	Form 1	Form 2	Form 3
22	56-86	44-76	43-79
21	49-82	36-71	38-74
20	41-77	29-67	30-69
19	34-72	22-62	25-62
18	28-67	17-58	21-55
17	22-61	14-51	17-48
16	18-56	11-44	14-43
15	15-49	9-36	11-38
14	12-41	6-29	8-30
13	10-34	5-22	6-25
12	8-28	3-17	4-21
11	5-22	2-14	4-17
10	4-18	1-11	3-14
9	3-15	1- 9	2-11
8	2-12	1- 6	1- 8
7	2-10	0- 5	0- 6
6	1- 8	0- 3	0- 4
5	0- 5	0- 2	0- 4
4	0- 4	0- 1	0- 3
3	0- 3	0- 1	0- 2
2	0- 2	0- 1	0- 1
1	0- 2	--	--
0	0- 1	--	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 26

Raw Scores and Corresponding Percentile Bands for the Gates-MacGinitie Reading Test, Survey E Comprehension, Forms 1, 2, and 3

Obtained Raw Score	Percentile Bands		
	Form 1	Form 2	Form 3
51	98-99+	99-99+	99-99+
50	97-99+	97-99+	98-99+
49	96-99+	96-99+	96-99+
48	93-99+	95-99+	94-99+
47	91-99	93-99	92-99+
46	89-98	90-99	89-99+
45	87-98	87-99	87-99
44	84-97	85-97	84-98
43	81-96	81-96	81-96
42	77-93	78-95	77-94
41	74-91	75-93	74-92
40	71-89	73-90	71-89
39	68-87	72-87	67-87
38	64-84	69-85	63-84
37	60-81	65-81	60-81
36	57-77	62-78	57-77
35	54-74	59-75	53-74
34	50-71	57-73	48-71
33	46-68	54-72	45-67
32	43-64	51-69	42-63
31	40-60	47-65	39-60
30	37-57	43-62	35-57
29	34-54	40-59	32-53
28	31-50	36-57	31-48
27	28-46	32-54	30-45
26	25-43	29-51	28-42
25	22-40	26-47	25-39

Table 26 (Continued)

Obtained Raw Score	Percentile Bands		
	Form 1	Form 2	Form 3
24	20-37	23-43	23-35
23	17-34	20-40	21-32
22	15-31	17-36	18-31
21	13-28	14-32	16-30
20	11-25	12-29	14-28
19	10-22	10-26	12-25
18	9-20	8-23	10-23
17	7-17	7-20	8-21
16	5-15	6-17	6-18
15	4-13	5-14	5-16
14	4-11	3-12	4-14
13	2-10	2-10	3-12
12	2- 9	1- 8	2-10
11	1- 7	1- 7	2- 8
10	1- 5	0- 6	1- 6
9	1- 4	0- 5	1- 5
8	0- 4	0- 3	1- 4
7	0- 2	0- 2	1- 3
6	0- 2	0- 1	0- 2
5	0- 1	0- 1	0- 2
4	0- 1	--	0- 1
3	0- 1	--	0- 1
2	--	--	0- 1
1	--	--	0- 1
0	--	--	--

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 27

Distribution of Standard z-Scores, T-Scores, and Stanines for the Gates-MacGinitie Reading Test, Survey E Speed and Accuracy, Form 1, Number Attempted

Raw Score[a]	z-Score	T-Score	Stanine
36	5.585	106	9
(35)[b]	5.349	104	
(34)	5.113	101	
(33)	4.877	99	
(32)	4.642	96	
(31)	4.406	94	
(30)	4.170	92	
(29)	3.934	89	
28	3.698	87	
(27)	3.462	85	
26	3.226	82	
25	2.991	80	
24	2.755	78	
23	2.519	75	
22	2.283	73	
21	2.047	71	
20	1.811	68	
19	1.576	66	8
18	1.340	63	
17	1.104	61	7
16	.867	59	

Table 27 (Continued)

Raw Score[a]	z-Score	T-Score	Stanine
15	.632	56	6
14	.396	54	
13	.160	52	
12	- .076	49	5
11	- .311	47	
10	- .547	45	4
9	- .783	42	
8	-1.019	40	3
7	-1.255	38	
6	-1.491	35	2
5	-1.726	33	
4	-1.962	30	1
3	-2.198	28	
(2)	-2.434	26	
1	-2.670	23	

[a]$n = 511$; $\bar{X} = 12.32$; $\sigma = 4.24$; range of obtained scores: 1-36.

[b]Numbers in parentheses represent interpolated scores.

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 28

Distribution of Standard z-Scores, T-Scores, and Stanines for the Gates-MacGinitie Reading Test, Survey E Speed and Accuracy, Form 2, Number Attempted

Raw Score[a]	z-Score	T-Score	Stanine
36	5.753	108	9
(35)[b]	5.509	105	
34	5.264	103	
(33)	5.020	100	
(32)	4.775	98	
(31)	4.531	95	
30	4.286	93	
(29)	4.042	90	
(28)	3.797	88	
(27)	3.553	86	
(26)	3.308	83	
25	3.064	81	
24	2.819	78	
23	2.575	76	
22	2.330	73	
21	2.086	71	
20	1.841	68	
19	1.597	66	8
18	1.352	64	
17	1.108	61	

Table 28 (Continued)

Raw Score[a]	z-Score	T-Score	Stanine
16	.863	59	7
15	.619	56	
14	.374	54	6
13	.130	51	
12	- .115	49	5
11	- .359	46	
10	- .604	44	4
9	- .848	42	
8	-1.093	39	3
7	-1.337	37	
6	-1.582	34	2
5	-1.826	32	
4	-2.071	29	1
3	-2.315	27	
(2)	-2.560	24	
(1)	-2.804	22	

[a] $n = 497$; $\bar{X} = 12.47$; $\sigma = 4.09$; range of obtained scores: 3-36.

[b] Numbers in parentheses represent interpolated or extrapolated scores.

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 29

Distribution of Standard z-Scores, T-Scores, and Stanines for the Gates-MacGinitie Reading Test, Survey E Speed and Accuracy, Form 3, Number Attempted

Raw Score[a]	z-Score	T-Score	Stanine
36	5.829	108	9
(35)[b]	5.581	106	
(34)	5.333	103	
33	5.087	101	
(32)	4.836	98	
(31)	4.588	96	
(30)	4.340	93	
(29)	4.092	91	
28	3.844	88	
27	3.596	86	
(26)	3.347	84	
25	3.099	81	
24	2.851	79	
23	2.603	76	
22	2.355	74	
21	2.107	71	
20	1.859	69	
19	1.610	66	8
18	1.362	64	
17	1.114	61	7
16	.866	59	

Table 29 (Continued)

Raw Score[a]	z-Score	T-Score	Stanine
15	.618	56	6
14	.370	54	
13	.122	51	5
12	- .127	49	
11	- .375	46	
10	- .623	44	4
9	- .871	41	
8	-1.119	39	3
7	-1.367	36	
6	-1.615	34	2
5	-1.864	31	
4	-2.112	29	1
(3)	-2.360	26	
(2)	-2.608	24	
(1)	-2.856	21	

[a] $n = 468$; $\bar{X} = 12.51$; $\sigma = 4.03$; range of obtained scores: 4-36.

[b] Numbers in parentheses represent interpolated or extrapolated scores.

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 30

Distribution of Standard z-Scores, T-Scores, and Stanines for the Gates-MacGinitie Reading Test, Survey E Speed and Accuracy, Form 1, Number Correct

Raw Score[a]	z-Score	T-Score	Stanine
(36)[b]	14.298	193	9
(35)	13.823	188	
(34)	13.349	184	
(33)	12.874	179	
(32)	12.399	174	
(31)	11.924	169	
(30)	11.449	165	
(29)	10.974	160	
(28)	10.500	155	
(27)	10.025	150	
(26)	9.550	146	
(25)	9.075	141	
(24)	8.600	136	
(23)	8.125	131	
(22)	7.651	127	
(21)	7.176	122	
(20)	6.701	117	
(19)	6.226	112	
(18)	5.751	108	
(17)	5.276	103	

Table 30 (Continued)

Raw Score[a]	z-Score	T-Score	Stanine
(16)	4.802	98	9
(15)	4.327	93	
(14)	3.852	89	
13	3.377	84	
12	2.902	79	
11	2.427	74	
10	1.953	70	
9	1.478	65	8
8	1.003	60	7
7	.528	55	6
6	.053	51	5
5	- .422	46	
4	- .897	41	4
3	-1.371	36	3
2	-1.846	32	2
1	-2.321	27	1
0	-2.796	22	

[a] $n = 511$; $\bar{X} = 5.888$; $\sigma = 2.106$; range of obtained scores: 0-13.

[b] Numbers in parentheses represent extrapolated scores.

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 31

Distribution of Standard z-Scores, T-Scores, and Stanines for the Gates-MacGinitie Reading Test, Survey E Speed and Accuracy, Form 2, Number Correct

Raw Score[a]	z-Score	T-Score	Stanine
(36)[b]	14.228	192	9
(35)	13.753	188	
(34)	13.278	183	
(33)	12.803	178	
(32)	12.329	173	
(31)	11.854	169	
(30)	11.379	164	
(29)	10.904	159	
(28)	10.429	154	
(27)	9.954	150	
(26)	9.480	145	
(25)	9.005	140	
(24)	8.530	135	
(23)	8.055	131	
(22)	7.580	126	
(21)	7.105	121	
(20)	6.631	116	
(19)	6.156	112	
(18)	5.681	107	
(17)	5.206	102	

Table 31 (Continued)

Raw Score[a]	z-Score	T-Score	Stanine
(16)	4.731	97	9
(15)	4.256	93	
(14)	3.782	88	
13	3.307	83	
12	2.832	78	
11	2.357	74	
10	1.882	69	
9	1.407	64	8
8	.933	59	7
7	.458	55	6
6	- .017	50	5
5	- .492	45	
4	- .967	40	4
3	-1.442	36	3
2	-1.917	31	2
1	-2.392	26	1
0	-2.867	21	

[a]$n = 497$; $\bar{X} = 6.036$; $\sigma = 2.106$; range of obtained raw scores: 0-13.

[b]Numbers in parentheses represent extrapolated scores.

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 32

Distribution of Standard z-Scores, T-Scores, and Stanines for the Gates-MacGinitie Reading Test, Survey E Speed and Accuracy, Form 3, Number Correct

Raw Score[a]	z-Score	T-Score	Stanine
(36)[b]	14.258	193	9
(35)	13.784	188	
(34)	13.309	183	
(33)	12.834	178	
(32)	12.359	174	
(31)	11.884	169	
(30)	11.409	164	
(29)	10.935	159	
(28)	10.460	155	
(27)	9.985	150	
(26)	9.510	145	
(25)	9.035	140	
(24)	8.560	136	
(23)	8.086	131	
(22)	7.611	126	
(21)	7.136	121	
(20)	6.661	117	
(19)	6.186	112	
(18)	5.711	107	
(17)	5.237	102	

Table 32 (Continued)

Raw Score[a]	z-Score	T-Score	Stanine
16	4.760	98	9
15	4.286	93	
14	3.811	88	
13	3.336	83	
12	2.862	79	
11	2.387	74	
10	1.912	69	
9	1.437	64	8
8	.963	60	7
7	.488	55	6
6	.013	50	5
5	- .462	45	
4	- .936	41	4
3	-1.411	36	3
2	-1.886	31	2
1	-2.360	26	1
(0)	-2.836	22	

[a]$n = 468$; $\bar{X} = 5.972$; $\sigma = 2.106$; range of obtained raw scores: 1-16.

[b]Numbers in parentheses represent extrapolated scores.

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 33

Distribution of Standard *z*-Scores, *T*-Scores, and Stanines for the Gates-MacGinitie Reading Test, Survey E Vocabulary Subtest, Form 1

Raw Score[a]	*z*-Score	*T*-Score	Stanine
(50)[b]	4.923	99	9
(49)	4.767	98	
(48)	4.611	96	
(47)	4.455	95	
(46)	4.299	93	
(45)	4.143	91	
(44)	3.987	90	
(43)	3.831	88	
(42)	3.675	87	
(41)	3.520	85	
(40)	3.364	84	
39	3.207	82	
38	3.052	81	
37	2.896	79	
(36)	2.740	77	
35	2.584	76	
34	2.428	74	
33	2.272	73	
32	2.116	71	
31	1.960	70	
30	1.804	68	
29	1.648	67	8
28	1.493	65	
27	1.337	63	
26	1.181	62	
25	1.025	60	7
24	.869	59	
23	.713	57	

Table 33 (Continued)

Raw Score[a]	z-Score	T-Score	Stanine
22	.557	56	6
21	.401	54	
20	.245	53	
19	.089	51	5
18	- .066	49	
17	- .222	48	
16	- .378	46	4
15	- .534	45	
14	- .690	43	
13	- .846	42	3
12	-1.002	40	
11	-1.158	38	
10	-1.314	37	2
9	-1.470	35	
8	-1.625	34	
7	-1.781	32	1
6	-1.937	31	
5	-2.093	29	
4	-2.249	28	
3	-2.405	26	
2	-2.561	24	
1	-2.717	23	
0	-2.873	21	

[a]$n = 509$; $\bar{X} = 18.426$; $\sigma = 6.414$; range of obtained raw scores: 0-39.

[b]Numbers in parentheses represent extrapolated or interpolated scores.

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 34

Distribution of Standard z-Scores, T-Scores, and Stanines for the Gates-MacGinitie Reading Test, Survey E Vocabulary Subtest, Form 2

Raw Score[a]	z-Score	T-Score	Stanine
(50)[b]	4.460	95	9
(49)	4.308	93	
(48)	4.157	92	
(47)	4.005	90	
(46)	3.854	89	
(45)	3.702	87	
(44)	3.550	86	
43	3.399	84	
42	3.247	83	
(41)	3.095	81	
40	2.944	79	
(39)	2.792	78	
38	2.640	76	
37	2.489	75	
36	2.337	73	
35	2.185	72	
34	2.034	70	
33	1.882	69	
32	1.731	67	
31	1.579	66	8
30	1.427	64	
29	1.276	63	
28	1.124	61	7
27	.973	60	
26	.821	58	
25	.669	57	6
24	.518	55	
23	.366	54	

Table 34 (Continued)

Raw Score[a]	z-Score	T-Score	Stanine
22	.214	52	6
21	.063	51	5
20	- .089	49	
19	- .240	48	
18	- .392	46	4
17	- .544	45	
16	- .695	43	
15	- .847	42	3
14	- .999	40	
13	-1.150	39	
12	-1.302	37	
11	-1.453	36	2
10	-1.605	34	
9	-1.757	32	
8	-1.908	31	1
(7)	-2.060	29	
6	-2.212	28	
5	-2.363	26	
4	-2.515	25	
3	-2.666	23	
(2)	-2.818	22	
(1)	-2.970	20	
0	-3.121	19	

[a] $n = 495$; $\bar{X} = 20.586$; $\sigma = 6.595$; range of obtained raw scores: 0-43.

[b] Numbers in parentheses represent extrapolated or interpolates scores.

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 35

Distribution of Standard z-Scores, T-Scores, and Stanines for the Gates-MacGinitie Reading Test, Survey E Vocabulary Subtest, Form 3

Raw Score[a]	z-Score	T-Score	Stanine
(50)[b]	4.696	97	9
(49)	4.539	95	
(48)	4.382	94	
(47)	4.226	92	
(46)	4.069	91	
(45)	3.912	89	
(44)	3.755	88	
(43)	3.598	86	
(42)	3.441	84	
(41)	3.284	83	
(40)	3.127	81	
(39)	2.970	80	
(38)	2.813	78	
37	2.656	77	
(36)	2.500	75	
35	2.343	73	
34	2.186	72	
33	2.029	70	
32	1.872	69	
31	1.715	67	
30	1.558	66	8
29	1.401	64	
28	1.244	62	
27	1.087	61	7
26	.930	59	
25	.773	58	
24	.617	56	6
23	.460	55	

Table 35 (Continued)

Raw Score[a]	*z*-Score	*T*-Score	Stanine
22	.303	53	6
21	.146	52	5
20	- .011	50	
19	- .168	48	
18	- .325	47	4
17	- .482	45	
16	- .639	44	
15	- .796	42	
14	- .953	41	3
13	-1.110	39	
12	-1.267	37	
11	-1.423	36	2
10	-1.580	34	
9	-1.737	33	
8	-1.894	31	1
7	-2.051	30	
6	-2.208	28	
5	-2.365	26	
4	-2.522	25	
(3)	-2.679	23	
2	-2.836	22	
(1)	-2.993	20	
(0)	-3.149	19	

[a] $n = 464$; $\bar{X} = 20.071$; $\sigma = 6.373$; range of obtained raw scores: 2-37.

[b] Numbers in parentheses represent extrapolated or interpolated scores.

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 36

Distribution of Standard z-Scores, T-Scores, and Stanines for the Gates-MacGinitie Reading Test, Survey E Comprehension Subtest, Form 1

Raw Score[a]	z-Score	T-Score	Stanine
(52)[b]	2.125	71	9
51	2.026	70	
50	1.928	69	
49	1.830	68	
48	1.732	67	
47	1.634	66	8
46	1.536	65	
45	1.437	64	
44	1.339	63	
43	1.241	62	
42	1.143	61	7
41	1.045	61	
40	.946	60	
39	.848	59	
38	.750	58	
37	.652	57	6
36	.554	56	
35	.456	55	
34	.357	54	
33	.259	53	
32	.161	52	5
31	.063	51	
30	- .035	50	
29	- .133	49	
28	- .232	48	
27	- .330	47	4
26	- .428	46	
25	- .526	45	

Table 36 (Continued)

Raw Score[a]	z-Score	T-Score	Stanine
24	- .624	44	4
23	- .722	43	
22	- .821	42	3
21	- .919	41	
20	-1.017	40	
19	-1.115	39	
18	-1.213	38	
17	-1.311	37	2
16	-1.410	36	
15	-1.508	35	
14	-1.606	34	
13	-1.704	33	
12	-1.802	32	1
11	-1.900	31	
10	-1.999	30	
9	-2.097	29	
(8)	-2.195	28	
7	-2.293	27	
6	-2.391	26	
5	-2.490	25	
(4)	-2.588	24	
(3)	-2.686	23	
2	-2.784	22	
(1)	-2.882	21	
(0)	-2.981	20	

[a] $n = 510$; $\bar{X} = 30.359$; $\sigma = 10.186$; range of obtained raw scores: 2-51.

[b] Numbers in parentheses represent interpolated or extrapolated scores.

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 37

Distribution of Standard z-Scores, T-Scores, and Stanines for the Gates-MacGinitie Reading Test, Survey E Comprehension Subtest, Form 2

Raw Score[a]	z-Score	T-Score	Stanine
(52)[b]	2.247	73	9
51	2.148	72	
50	2.049	71	
49	1.950	70	
48	1.851	69	
47	1.752	68	
46	1.653	67	8
45	1.554	66	
44	1.455	65	
43	1.356	64	
42	1.257	63	
41	1.158	62	7
40	1.059	61	
39	.960	60	
38	.861	59	
37	.762	58	
36	.664	57	6
35	.565	56	
34	.466	55	
33	.367	54	
32	.268	53	
31	.169	52	5
30	.070	51	
29	- .029	50	
28	- .128	49	
27	- .227	48	
26	- .326	47	4
25	- .425	46	

Table 37 (Continued)

Raw Score[a]	z-Score	T-Score	Stanine
24	- .524	45	4
23	- .623	44	
22	- .722	43	
21	- .821	42	3
20	- .920	41	
19	-1.019	40	
18	-1.118	39	
17	-1.217	38	
16	-1.315	37	2
15	-1.414	36	
14	-1.513	35	
13	-1.612	34	
12	-1.711	33	
11	-1.810	32	1
10	-1.909	31	
9	-2.008	30	
8	-2.107	29	
7	-2.206	28	
6	-2.305	27	
(5)	-2.404	26	
(4)	-2.503	25	
(3)	-2.602	24	
(2)	-2.701	23	
(1)	-2.800	22	
(0)	-2.899	21	

[a] $n = 496$; $\bar{X} = 29.294$; $\sigma = 10.106$; range of obtained raw scores: 6-51.

[b] Numbers in parentheses represent extrapolated scores.

Note. 1977 norms, Putnam City Schools, grade 6.5.

Table 38

Distribution of Standard z-Scores, T-Scores, and Stanines for the Gates-MacGinitie Reading Test, Survey E Comprehension Subtest, Form 3

Raw Score[a]	z-Score	T-Score	Stanine
(52)[b]	2.101	71	9
51	2.005	70	
(50)	1.909	69	
49	1.813	68	
48	1.717	67	8
47	1.622	66	
46	1.526	65	
45	1.430	64	
44	1.334	63	
43	1.239	62	
42	1.143	61	7
41	1.047	61	
40	.951	60	
39	.855	59	
38	.760	58	
37	.664	57	6
36	.568	56	
35	.472	55	
34	.376	54	
33	.281	53	
32	.185	52	5
31	.089	51	
30	- .007	50	
29	- .102	49	
28	- .198	48	
27	- .294	47	4
26	- .390	46	
25	- .486	45	

Table 38 (Continued)

Raw Score[a]	z-Score	T-Score	Stanine
24	- .581	44	4
23	- .667	43	
22	- .773	42	
21	- .869	41	3
20	- .965	40	
19	-1.060	39	
18	-1.156	38	
17	-1.252	38	
16	-1.348	37	2
15	-1.444	36	
14	-1.539	35	
13	-1.635	34	
12	-1.731	33	
11	-1.827	32	1
10	-1.922	31	
9	-2.018	30	
8	-2.114	29	
7	-2.210	28	
6	-2.306	27	
(5)	-2.401	26	
(4)	-2.500	25	
3	-2.593	24	
(2)	-2.689	23	
(1)	-2.785	22	
0	-2.880	21	

[a] $n = 471$; $\bar{X} = 30.070$; $\sigma = 10.440$; range of obtained raw scores: 0-51.

[b] Numbers in parentheses represent extrapolated or interpolated scores.

Note. 1977 norms, Putnam City Schools, grade 6.5.

A Comparison of Norms

Tables 39-42 exhibit the standard score distributions (T-scores) obtained from the administrations of the Gates-MacGinitie Reading Test, Survey E, Forms 1, 2, and 3 to three groups of varying abilities. Presented for comparison are results of the reading test performance of (a) the 1977 Putnam City norm group assessed at grade placement 6.5; (b) the 1965 Metropolitan School Study Council (MSSC) norm group assessed at grade placement 6.5; and (c) that portion of the original norm group assessed at grade placement 7.1 in 1964.

The standard score norms resulting from the present study are presented separately for Forms 1, 2, and 3 of each subtest. The MSSC and original (national) norms, presented in Tables 39 through 42 as they appeared in the original studies, represent results of the combined forms of each subtest. The conclusions drawn from comparisons of these norm scales are presented in Chapter V.

Table 39

A Comparison of Standard Score Distributions for Groups Administered the Gates-MacGinitie Reading Test, Survey E, Speed and Accuracy, Number Attempted

Obtained Raw Scores	Distribution of Standard Scores (T-Scores)[a]				
	Putnam City Group Norms for Grade Placement 6.5			MSSC Norms for G.P.[b] 6.5	Original Norms for G.P.[b] 7.1
	Form 1	Form 2	Form 3	Combined Test Forms	Combined Test Forms
36	106	108	108	73	--
35	--	--	--	72	--
34	--	103	--	71	--
33	--	--	101	71	--
32	--	--	--	70	--
31	--	--	--	69	--
30	--	93	--	68	--
29	--	--	--	68	--
28	87	--	88	67	75
27	85	--	86	65	73
26	82	--	--	64	71
25	80	81	81	63	70
24	78	78	79	62	69
23	75	76	76	60	68
22	73	73	74	59	67
21	71	71	71	57	66
20	68	68	69	56	65

Table 39 (Continued)

Obtained Raw Scores	Distribution of Standard Scores (T-Scores)[a]				
	Putnam City Group Norms for Grade Placement 6.5			MSSC Norms for G.P.[b] 6.5	Original Norms for G.P.[b] 7.1
	Form 1	Form 2	Form 3	Combined Test Forms	Combined Test Forms
19	66	66	66	54	64
18	63	64	64	52	62
17	61	61	61	50	60
16	59	59	59	49	58
15	56	56	56	47	56
14	54	54	54	45	54
13	52	51	51	42	51
12	49	49	49	40	49
11	47	46	46	38	47
10	45	44	44	36	44
9	42	42	41	34	41
8	40	39	39	31	38
7	38	37	36	29	35
6	35	34	34	--	31
5	33	32	31	--	--
4	30	29	29	--	--
3	28	27	--	--	--
2	--	--	--	--	--
1	23	--	--	--	--

[a] $\underline{X} = 50$; $\sigma = 10$

[b] G.P.: Grade Placement

Table 40

A Comparison of Standard Score Distributions for Groups Administered the Gates-MacGinitie Reading Test, Survey E, Speed and Accuracy; Number Correct

Obtained Raw Scores	Distribution of Standard Scores (T-Scores)[a]				
	Putnam City Group Norms for Grade Placement 6.5			MSSC Norms for G.P.[b] 6.5	Original Norms for G.P.[b] 7.1
	Form 1	Form 2	Form 3	Combined Test Forms	Combined Test Forms
30	--	--	--	75	--
29	--	--	--	73	--
28	--	--	--	71	--
27	--	--	--	70	--
26	--	--	--	69	--
25	--	--	--	68	--
24	--	--	--	66	--
23	--	--	--	65	--
22	--	--	--	63	76
21	--	--	--	61	74
20	--	--	--	59	72
19	--	--	--	57	70
18	--	--	--	55	68
17	--	--	--	53	66
16	--	--	98	50	64
15	--	--	93	49	61
14	--	--	88	47	59
13	84	83	83	45	56

Table 40 (Continued)

Obtained Raw Scores	Distribution of Standard Scores (T-Scores)[a]				
	Putnam City Group Norms for Grade Placement 6.5			MSSC Norms for G.P.[b] 6.5	Original Norms for G.P.[b] 7.1
	Form 1	Form 2	Form 3	Combined Test Forms	Combined Test Forms
12	79	78	79	42	54
11	74	74	74	41	51
10	70	69	69	38	49
9	65	64	64	36	46
8	60	59	60	34	44
7	55	55	55	32	41
6	51	50	50	30	38
5	46	45	45	28	36
4	41	40	41	--	33
3	36	36	36	--	31
2	32	31	31	--	--
1	27	26	26	--	--
0	22	21	--	--	--

[a] $\bar{X} = 50$; $\sigma = 10$

[b] G.P.: Grade Placement

Table 41

A Comparison of Standard Score Distributions for Groups Administered the Gates-MacGinitie Reading Test, Survey E, Vocabulary Subtest

Obtained Raw Scores	Distribution of Standard Scores (T-Scores)[a]				
	Putnam City Group Norms for Grade Placement 6.5			MSSC Norms for G.P.[b] 6.5	Original Norms for G.P.[b] 7.1
	Form 1	Form 2	Form 3	Combined Test Forms	Combined Test Forms
43	--	84	--	--	--
42	--	83	--	--	--
41	--	--	--	--	--
40	--	79	--	75	75
39	82	--	--	73	74
38	81	76	--	72	72
37	79	75	77	71	71
36	--	73	--	69	70
35	76	72	73	68	68
34	74	70	72	66	67
33	73	69	70	65	66
32	71	67	69	63	65
31	70	66	67	62	64
30	68	64	66	60	63
29	67	63	64	59	62
28	65	61	62	57	60
27	63	60	61	56	59
26	62	58	59	54	58
25	60	57	58	53	57
24	59	55	56	51	55
23	57	54	55	50	54
22	56	52	53	49	53
21	54	51	52	47	52

Table 41 (Continued)

Obtained Raw Scores	Distribution of Standard Scores (T-Scores)[a]				
	Putnam City Group Norms for Grade Placement 6.5			MSSC Norms for G.P.[b] 6.5	Original Norms for G.P.[b] 7.1
	Form 1	Form 2	Form 3	Combined Test Forms	Combined Test Forms
20	53	49	50	46	51
19	51	48	48	44	49
18	49	46	47	43	48
17	48	45	45	41	47
16	46	43	44	40	45
15	45	42	42	39	44
14	43	40	41	37	42
13	42	39	39	36	41
12	40	37	37	34	39
11	38	36	36	33	38
10	37	34	34	31	36
9	35	32	33	30	34
8	34	31	31	28	32
7	32	--	30	27	30
6	31	28	28	--	--
5	29	26	26	--	--
4	28	25	25	--	--
3	26	23	--	--	--
2	24	--	22	--	--
1	23	--	--	--	--
0	21	19	--	--	--

[a] $\bar{X} = 50$; $\sigma = 10$

[b] G.P.: Grade Placement

Table 42

A Comparison of Standard Score Distributions for Groups Administered the Gates-MacGinitie Reading Test, Survey E, Comprehension Subtest

Obtained Raw Scores	Distribution of Standard Scores (T-Scores)[a]				
	Putnam City Group Norms for Grade Placement 6.5			MSSC Norms for G.P.[b] 6.5	Original Norms for G.P.[b] 7.1
	Form 1	Form 2	Form 3	Combined Test Forms	Combined Test Forms
52	--	--	--	--	--
51	70	72	70	76	--
50	69	71	--	72	75
49	68	70	68	70	72
48	67	69	67	68	69
47	66	68	66	66	67
46	65	67	65	64	65
45	64	66	64	62	64
44	63	65	63	60	63
43	62	64	62	59	61
42	61	63	61	58	60
41	61	62	61	57	59
40	60	61	60	55	58
39	59	60	59	54	57
38	58	59	58	53	56
37	57	58	57	52	55
36	56	57	56	51	55
35	55	56	55	50	54

Table 42 (Continued)

Obtained Raw Scores	Distribution of Standard Scores (T-Scores)[a]				
	Putnam City Group Norms for Grade Placement 6.5			MSSC Norms. for G.P.[b] 6.5	Original Norms for G.P.[b] 7.1
	Form 1	Form 2	Form 3	Combined Test Forms	Combined Test Forms
34	54	55	54	50	53
33	53	54	53	49	52
32	52	53	52	48	52
31	51	52	51	47	51
30	50	51	50	47	50
29	49	50	49	46	50
28	48	49	48	45	49
27	47	48	47	44	48
26	46	47	46	44	47
25	45	46	45	43	47
24	44	45	44	42	46
23	43	44	43	41	45
22	42	43	42	41	45
21	41	42	41	40	44
20	40	41	40	39	43
19	39	40	39	38	42
18	38	39	38	37	42
17	37	38	38	36	41
16	36	37	37	35	40
15	35	36	36	34	39

Table 42 (Continued)

Obtained Raw Scores	Distribution of Standard Scores (T-Scores)[a]				
	Putnam City Group Norms for Grade Placement 6.5			MSSC Norms for G.P.[b] 6.5	Original Norms for G.P.[b] 7.1
	Form 1	Form 2	Form 3	Combined Test Forms	Combined Test Forms
14	34	35	35	34	38
13	33	34	34	33	37
12	32	33	33	32	36
11	31	32	32	30	35
10	30	31	31	29	34
9	29	30	30	28	32
8	--	29	29	27	31
7	27	28	28	--	29
6	26	27	27	--	--
5	25	--	--	--	--
4	--	--	--	--	--
3	--	--	24	--	--
2	22	--	--	--	--
1	--	--	--	--	--
0	--	--	21	--	--

[a] $\bar{X} = 50$; $\alpha = 10$

[b] G.P.: Grade Placement

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The problem as stated in the present study was to establish local sixth-grade norms for the Gates-MacGinitie Reading Test, Survey E, Forms 1, 2, and 3. This test, originally constructed and normed for use in grades 7, 8, and 9, consists of Speed and Accuracy, Vocabulary, and Comprehension subtests.

The students for which these special-study norms were developed attend 13 elementary schools in Putnam City District, a predominantly middle-class suburban public school district of 20,000 students located in the northwest quadrant of Oklahoma City, Oklahoma. The standardized test results presented in Table 2 reveal district grade level means of academic aptitude scores which are greater than the national mean to a statistically significant degree. The standardized reading test results presented in Table 1 reveal that these students achieve at levels which are not only well above their grade placement, but also above their anticipated achievement, based on measures of their academic aptitude.

The need for the present study was to provide a means of obtaining appropriate test data to use with students' grade point averages and teachers' recommendations to place Putnam City sixth-grade students in ability-grouped seventh-grade reading classes. The purpose was therefore (a) to determine which of the Gates-MacGinitie Reading Tests could be appropriately adapted for use with these students; (b) to administer the test; and (c) to provide norm scales with which the obtained raw scores could be interpreted most accurately.

Both Surveys D and E of the Gates-MacGinitie Reading Test series were administered during the 1973-76 pilot projects conducted for the present study. The results indicated that the ceiling and norms for Survey D (nationally normed for grades 4, 5, and 6) were too low to provide for accurate interpretation of obtained raw scores. Although the ceiling for Survey E (intended for grades 7, 8, and 9) appeared to be adequate, it was theorized that use of the national norms for this test might lead to errant estimates of students' reading achievement levels.

A review of the literature revealed that, for groups whose mean achievement levels deviate sufficiently on either side of the national mean to exempt them as representative samples of the national norm group, the use of local or special-study norms provides for a more appropriate interpretation of obtained raw scores. Among the many examples of

the specialized norm studies reviewed, the most relevant for the present study was the establishment of Gates-MacGinitie Reading Test norms for use in the 70 suburban New York districts comprising the Metropolitan School Study Council (MSSC). According to the authors of this series of reading tests, the superior academic performance of the MSSC students necessitated the development of special-study norms to more adequately reflect these students' reading achievement levels. The similarity in test performance of the Putnam City and MSSC students led to questions regarding the usefulness of MSSC norms for the local sixth-grade population.

Ten null hypotheses were formulated and tested in order to determine and/or establish (a) the need for specialized norms for this population; (b) the equivalence of academic aptitude among the three samples tested; and (c) the equivalence of the three test forms in measuring reading achievement. Differences in the following measures were tested for statistical significance at the .05 level:

1. Means of Total IQ scores among the three local norm samples represented by Forms 1, 2, and 3 of Survey E (Ho_1);

2. Means of obtained raw scores among the three local norm groups in regard to each of the subtests administered (Ho_2 through Ho_4);

3. The Putnam City raw score means for each subtest as compared with those obtained for the seventh-grade portion of the original norm group (Ho_5 through Ho_7);

4. The Putnam City subtest raw score means at grade placement 6.5 compared with the interpolated raw score means equivalent to the grade equivalent of 6.5 on each of the original norm scales (Ho_8 through Ho_{10}).

The research design for the present study employed a form of cluster sampling from among the entire population of sixth-grade students (with the exception of those enrolled in classes for the Educable and Trainable Mentally Retarded). Students were divided into three subgroups (one for each test form) through a numbering-off system. This was accomplished in each classroom setting to achieve total random distribution of abilities across the three samples. The samples numbered from 486 to 511, for a total of 1,478 students drawn from the 1,495 eligible subjects.

Each student was administered all subtests of a given test form within a specified three-day limit in mid-February 1977. All tests were administered under normal classroom conditions by the students' regular reading teachers, 86% of whom were certified Reading Specialists. The test administrators were given both written and oral directions as well as test manuals, students' test booklets, and machine-scorable answer sheets during a pre-test instructional session.

The answer sheets were coded with students' IQ scores obtained from an October 1976 administration of the Short Form Test of Academic Aptitude (McGraw-Hill, 1970 ed.).

These forms were machine-scored during which process all relevant information regarding name, IQ score, and response to each test item was transferred to IBM cards. Computer programs were utilized to obtain the necessary descriptive statistics for testing the hypotheses and constructing the norm scales.

The results of testing the first four hypotheses with the F-test for the analysis of variance indicated that no statistically significant differences existed among the sample means of total IQ scores nor among the means of reading subtests for each form. Hypotheses 5, 6, and 7 were tested with the two-sample z-test for an hypothesis about a mean (a test for large samples with unequal variances). Comparisons of Putnam City group means for each subtest with those of the seventh-grade portion of the original norm group resulted in the following:

1. On all forms of the Speed and Accuracy subtests, the students from the original norm groups performed at levels which were higher to a statistically significant degree.

2. No statistically significant differences were found in performance on Forms 1 and 3 of the Vocabulary and Comprehension subtests.

3. Statistically significant differences were found in the raw score means for Form 2 of the Vocabulary and Comprehension subtests. The Putnam City sample means for these

subtests were found to be significantly greater than those obtained from the seventh-grade portion of the original norm group.

Hypotheses 8, 9, and 10 were tested with the one-sample z-test for an hypothesis about a mean to compare the raw score means of Putnam City students tested at grade placement 6.5 with the raw score equivalents of the interpolated grade equivalent score of 6.5 on the original norm scale for each subtest. Results of these tests indicated the following:

1. Statistically significant differences were found in the means being compared in regard to Speed and Accuracy (number correct). The Putnam City mean was significantly less than that estimated for grade placement 6.5 on the original norm scale for each test form.

2. Statistically significant differences were found in the means being compared in regard to both the Vocabulary and Comprehension subtests. In all instances, the Putnam City means were found to be significantly greater than the interpolated means of the original norm group.

The question regarding the comparison of MSSC students' performance levels at grade placement 6.5 and those of the Putnam City norm samples was addressed through application of the one-sample z-test for an hypothesis about a mean. Performance levels of the MSSC groups on all forms of each subtest were found to be consistently superior to those of the Putnam City groups to a statistically significant degree.

Utilizing the data gathered for the present study, norm scales were constructed for each of the three forms of each subtest. Separate scales were established for the Speed and Accuracy subtest, "number attempted" and "number correct." These norms were expressed in percentile ranks, percentile bands, z-scores, T-scores, and stanines. Grade equivalent scales were purposely omitted from those norms established by this study. However, the obtained sample mean for each subtest form can be considered the equivalent of 6.5 on a grade equivalent scale.

In order to facilitate placement of the 1977 and subsequent sixth-grade students in seventh-grade reading classes, students' raw and converted scores and the norm scales established in this study were presented to the appropriate junior high school counselors and to the appropriate sixth- and seventh-grade teachers. The presentation of these data was accompanied by an explanation of the meaning of each type of score and instructions for their use in the placement process.

The exclusive use of subjects from among the sixth-grade population of Putnam City School District places certain limitations on the use of the norms established in this study. These data should not be generalized to populations other than that utilized to establish these norms.

Conclusions

Based on the findings of this study, the following is concluded:

1. The Gates-MacGinitie Reading Test scores of Putnam City sixth-grade students will be more appropriate for placement purposes when derived from local norms than when they are derived from national or MSSC norms.

2. In regard to the Speed and Accuracy subtests, use of national norms to interpret Putnam City students' test scores will lead to overestimates of reading achievement levels.

3. The equivalent distribution of academic ability and achievement levels across subgroups of the Putnam City population can be achieved through use of the sampling procedures employed in this study.

4. In regard to the Gates-MacGinitie Reading Test, Survey E, statistically equivalent test results (determined at the .05 level) can be obtained through the administration of any form of any given subtest to ability-equivalent subgroups of the Putnam City sixth-grade population.

5. The local norms established separately for each form of the respective subtests can be safely combined on any one of the subtests to facilitate use of these norms in Putnam City District.

6. Local norms in the form of grade equivalent scales can not be constructed from the available data with any degree of accuracy.

7. The Vocabulary and Comprehension subtests administered in this study can be considered as reliable instruments for this population. However, there is insufficient evidence to consider the Speed and Accuracy subtests as reliable instruments for measuring reading abilities of Putnam City sixth-grade students.

8. The respective Putnam City teachers and counselors for whom these data were made available can adapt to the use of standard scores and percentile ranks in lieu of the grade equivalent scores formerly required for placement of these students in ability-grouped seventh-grade reading classes.

Recommendations and Discussion

Based on the findings of this study, the following recommendations are made:

1. That this placement testing program be continued, and the resulting information be distributed for use in the manner specified in this study.

2. That the Speed and Accuracy subtests be administered with an expanded time limit and new norms be established on the basis of the data gathered. The use of an expanded time limit would aid in determining whether the level of Putnam City students' performance resulting from the present study was attributable to the difficulty level of the test ceiling or to the length of time allowed for responding to the test items.

3. That a test-retest pattern be established in a future administration of the Speed and Accuracy subtests in order to provide more appropriate data with which to determine the reliability of these tests.

4. That the results of further standardized testing programs in Putnam City District reflect the use of both local and national norms. The publishers of the standardized tests administered yearly in this district will provide this service upon request.

5. That the Putnam City personnel who utilize students' test scores analyze and study the comparative interpretations of raw scores based on both local and national norms to maintain a realistic perspective in terms of these students' actual reading achievement levels.

6. That a means be provided whereby the results obtained from the administration of the Gates-MacGinitie Reading Test, Survey E, may be expressed in terms of estimated functional reading levels. To accomplish this task, students should be administered Survey E and an informal reading inventory within a period of time not to exceed two or three weeks. A correlation of the results will provide the basis for predicting functional reading levels based on the Gates-MacGinitie Reading Test results. The use of this procedure will further facilitate (a) the organization and scheduling of classes designed to meet instructional needs of students performing at varying reading levels; (b) the

appropriate placement of students in these classes; and (c) more precise placement of students in instructional materials designed to meet the needs of varying specified reading ability levels.

7. That relevant norms be established for all school populations whose grade level means of reading achievement scores deviate significantly either side of the national mean for groups tested at an equivalent grade level.

The speed and accuracy score distributions (number correct) were positively skewed and platykurtic. These test results, found to be consistent across all three forms, would be expected in a district such as Putnam City wherein, through the end of grade 6, phonics, structural analysis, vocabulary, and comprehension are stressed instead of speed. Evidence of such priorities placed on the instruction of these specific skills can be found in (a) a review of teachers' stated instructional objectives and (b) observation of classroom practices in the teaching of reading.

Treatment of the Speed and Accuracy subtest data from each of the three test forms therefore yielded results which could not satisfy the reliability criteria established by Guilford (1954, p. 392) and Gulliksen (1950, p. 235). However, the norms constructed in the present study are considered usable until different norms can be developed using the test-retest method to obtain additional data.

BIBLIOGRAPHY

Books

Angoff, W. H. Scales, norms, and equivalent scores. In R. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Brown, F. G. Principles of educational and psychological testing. Hinsdale, Ill.: The Dryden Press, Inc., 1970.

Burke, C. L. Tests and reviews: the GATES-MacGINITIE READING TESTS. In O. K. Buros (Ed.), The seventh mental measurements yearbook. Highland Park, N.J.: Gryphon, 1972.

Dixon, W. J. (Ed.). BMD biomedical computer programs. Berkeley: University of California Press, 1975.

Gellman, E. S. Statistics for teachers. New York: Harper & Row, 1973.

Guilford, J. P. Psychometric methods. New York: McGraw-Hill, 1974.

Gulliksen, H. Theory of mental tests. New York: John Wiley and Sons, Inc., 1950.

Kerlinger, F. N. Foundations of behavioral research. New York: Holt, Rinehart, and Winston, 1973.

Lindeman, R. H. Educational measurement. Glenview, Ill.: Scott, Foresman, and Co., 1967.

Mehrens, W. A., and Lehmann, I. J. Measurement and evaluation in education and psychology. New York: Holt, Rinehart, and Winston, 1973.

Schrader, W. B. Norms. In W. S. Monroe (Ed.), Encyclopedia of educational research (3rd ed.). New York: MacMillan, March 1958.

Thorndike, R. L., and Hagen, E. Measurement and evaluation in psychology and education. New York: John Wiley and Sons, Inc., 1969.

ERIC Documents

Bergsten, J. W. The effects of cluster sampling in the norming of an achievement test battery. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 25-March 1, 1973, pp. 1-8. (ERIC Document Reproduction Service No. ED 135 831)

Bradley, P. A., and Woolley, D. Making better decisions on assessed needs: Differentiated school norms. Los Angeles: Center for the Study of Evaluation, University of California at Los Angeles. (ERIC Document Reproduction Service No. ED 050 156)

Evard, B. L., and McGrady, H. J. Development of local language norms for Papago Indians, Mexican-Americans, Blacks, and Anglos. Paper presented at the Annual Meeting of the American Speech and Hearing Association, Las Vegas, 1974, pp. 3-15. (ERIC Document Reproduction Service No. ED 105 732)

Huyser, R., Fisher, T., and Olsen, J. Local and district school report: Explanatory materials (3rd report of the Michigan Assessment Program). Lansing, Michigan: State Department of Education, April 1973, pp. 4-37. (ERIC Document Reproduction Service No. ED 104 900)

Rasp, A., Jr., and Stiles, R. Using the ANCHOR TEST STUDY in state assessment. Paper presented at the Annual Meeting of the 60th American Research Association, San Francisco, April 19-23, 1976, pp. 2-9. (ERIC Document Reproduction Service No. ED 124 576)

Rasp, A., Jr. Using ANCHOR TEST STUDY tables in state assessment programs. Washington: U.S. Government Printing Office, 1975. (ERIC Clearinghouse on Tests, Measurement, and Evaluation; Document No. ED 135 842)

Schwartz, H. P. Testing for instructional purposes: Norm referenced--criterion referenced. Paper presented at the State Convention of the Kansas Personnel and Guidance Association, 1974, pp. 1-10. (ERIC Document Reproduction Service No. ED 096 335)

Waggemer, R., Starr, R. P., and Hopkins, T. Norming of a basic learning exam and its use to predict GED results. Paper presented at the Annual Meeting of the American Personnel and Guidance Association, New Orleans, April 1974, pp. 1-4. (ERIC Document Reproduction Service No. 096 324)

Periodicals

Alvord, D. J., and Brittingham, B. E. Evaluating performance on national assessment objectives: Norm-referenced and criterion-referenced interpretations. Journal of Educational Research, 1974, 68 (2), 59-61.

Cronbach, L. J., and Warrington, W. G. Time-limit tests: estimating their reliability and degree of speeding. Psychometrika, 1951, 16 (2), 167-189.

Broadley, G., and Broadley, K. M. Rural standardization of the BURT-VERNON GRADED WORD READING TEST. Alberta Journal of Educational Research, 1975, 21 (4), 289-294.

Lord, F. M. Test norms and sampling theory. Journal of Experimental Education, 1959, 27, 249-263.

Vernon, P. E. A restandardization of the BURT-VERNON GRADED WORD READING TEST. Western Psychologist, 1973, 4, 72-77.

Personal Communication

MacGinitie, Walter H. Personal interview, May 4, 1977, Miami Beach, Florida.

Olsen, June. Personal communication, September 1978.

Technical Reports

Adkins, E. R. ARMED SERVICES VOCATIONAL APTITUDE BATTERY, form 5: High school norming and standardization (final report no. AFHRL-TR-76-79). Annandale, Virginia: J. W. K. International Corporation, 1976, pp. 5-13.

American Psychological Corporation, Technical recommendations for psychological tests and diagnostic techniques. Psychological Bulletin, 51 (2), part 2, March 1954.

Department of Defense. ARMED SERVICES VOCATIONAL APTITUDE BATTERY counselor's guide (DOD 1304.12X), 1976.

Department of Defense. Technical research note 75-76: Representative normative tables for the ARMED SERVICES VOCATIONAL APTITUDE BATTERY (form 5), September 1976.

Durost, W. N., Balow, I. H., Bixler, H. H., Prescott, G. A., and Wrightstone, J. W. METROPOLITAN ACHIEVEMENT TESTS: Teacher's handbook. Dallas: Harcourt Brace Jovanovich, 1971.

Educational Testing Service. Cooperative school and college ability tests: Manual for interpreting scores. Princeton: E. T. S., 1957.

Gates, A. I., and MacGinitie, W. H. Metropolitan school study council norms for the GATES-MacGINITIE READING TESTS. New York: Teachers College Press, 1966.

Gates, A. I., and MacGinitie, W. H. Technical manual for the GATES-MacGINITIE READING TESTS. New York: Teachers College Press, 1972.

Lennon, R. T. Test service notebook no. 13: A glossary of 100 measurement terms. Dallas: Harcourt, Brace, and World (no date supplied).

Loret, P. G., Seder, A., Bianchini, J. C., and Vale, C. A. Anchor test study: Equivalence and norms tables for selected reading achievement tests (grades 4, 5, and 6). Washington, D.C.: U.S. Government Printing Office, 1975, pp. 1-11.

Mitchell, B. C. Test service notebook no. 13: A glossary of measurement terms. Dallas: Harcourt Brace Jovanovich (no date supplied).

Seashore, H. G., and Ricks, J. H., Jr. Norms must be relevant. Test Service Bulletin, nos. 36-40, 1948-1950, The Psychological Corporation, pp. 16-20.

Weshman, A. G. The three-legged coefficient (the meaning of correlations). Test Service Bulletin, nos. 36-40, 1948-1950, The Psychological Corporation, December 1950, pp. 20-22.

Unpublished

Pearson, M. A. The establishment of SCHOOL AND COLLEGE ABILITY TEST norms for blind children in grades 4, 5, and 6. Unpublished doctoral dissertation, University of Oklahoma, 1962.

Putnam City School District. Office of Curriculum and Instruction. Survey of number and location of reading specialists in Putnam City district, 1973-77.

Putnam City School District. Parent-Teacher Association. Annual membership report, 1977.

Putnam City School District. Personnel Office:
Membership Summary, January 1977;
Statistical Report, January 1977;
Personnel Policies Handbook, 1977.

APPENDIX

Instructions for Test Administrators

I. Preparation for Test Administration:

A. Give all students a "pep talk" to include the rationale for doing their very best on the test.

This is a special project designed to help place each student in appropriate reading and English classes for next year. Encourage them to take it seriously and see how well they can do.

B. Hand out one answer sheet to each child (for Speed and Accuracy subtest). Use the following steps to fill out the answer sheet: (See example).

1. School name
2. City name
3. Teacher's name
4. Grade 6.5
5. Print name in blocks according to directions on answer sheet: last, first, and middle initial. Have students fill in the appropriate spaces for the letters of their names.
6. For the first subtest, fill in "1" to indicate Speed and Accuracy. (At the subsequent test sessions, fill in #2 for Vocabulary, and #3 for Comprehension.)
7. Fill in "Spring Semester"
8. a. Using blocks 1 and 2, fill in the appropriate code (listed below) for your school.

Apollo. 01
Central Intermediate. 02
Coronado. 03
Harvest Hills 04
Hilldale. 05
Kirkland. 06
Lake Park 07
Overholser. 08
Rollingwood 09
Tulakes 10
Western Oaks. 11
Wiley Post. 12
Windsor Hills 13

b. Leave block #3 blank.

c. In blank #4, put the class #. (This class # should correspond with that used on your divider sheets in stacking the tests after administration).

d. Leave blanks #5 and #6 blank.

9. Fill in Grade (6).
10. Fill in birth date--Month and Year.
11. Fill in sex.
12. Have students "number off" orally "1, 2, 3; 1, 2, 3; etc." Tell them that these numbers are important, to remember them, and that they will be assigned these same numbers throughout the administration of the entire test (all subtests). This number indicates the form (E-1, E-2, or E-3) that each will be administered.
13. Have students #1 put an E-1 in the blank labeled Test. Students #2 will put E-2, and students #3 will put E-3.

End of answer sheet labeling.

14. Remind students of the following important factor: Use a #2 (dark) pencil and mark only the appropriate space for each answer.
15. When each subtest administration is completed, the easiest way to take them up is to make stacks of all E-1s together, E-2s together, etc.

II. Test Administration:

Using your Teacher's Manual for the Gates-MacGinitie Reading Test, administer the test according to the written directions. Read all student directions orally and use the examples provided. Question to make certain all students understand what they are to do.

III. Arrangement of Answer Sheets for Machine-Scoring:

In order to prepare the completed answer sheets for machine-scoring, please arrange in the following manner: The attached divider sheets are labeled and numbered in the correct order of arrangement for your convenience. Please do not bend, staple, or use rubber bands or paper clips with the answer sheets as they must be fed through a machine.

Arrangement of Answer Sheets:

E-1: Subtest I--Speed and Accuracy
Class #1 (arranged alphabetically by students' last names--boys first and then girls).

Class #2, Class #3, Class #4, etc., same arrangement by class as above.

E-1: Subtest II--Vocabulary

Class #1--Alphabetize, arrange by last name, boys first and then girls.

Class #2--Same as above.

E-1: Subtest III--Comprehension

Class #1
Class #2
Class #3

Do the same for each subtest and class for E-2, and E-3.

Please return all test forms and answer sheets by Friday, March 4, 1977. You may send these to my office through school mail.

Thank you.

Zoe Walker