

1954

A summary of research in test construction and evaluation

<https://hdl.handle.net/2144/9149>

Downloaded from DSpace Repository, DSpace Institution's institutional repository

Ed.
Thesis
Cullen, R. B.
1954
Stored

BOSTON UNIVERSITY
SCHOOL OF EDUCATION

Thesis
A Summary of Research in
Test Construction and Evaluation

Submitted by
Ruth B. Cullen
(B.S. in Educ., Lowell State Teachers College 1951)
In partial Fulfillment of Requirements for the Degree of
Master of Education

1954

Boston University
School of Education
Library

ACKNOWLEDGEMENT

Grateful acknowledgement is made for the help and guidance of our advisor, Dr. Donald D. Durrell.

predicting first grade success was the primary factor for investigation in several studies. Berwick (4) using the Kuhlman Anderson Intelligence Tests, the Lee Clark Reading Readiness Tests and Gates Primary Reading Tests on 238 children in a community of greater Boston, concluded that all three instruments had about the same predictive value., (13% better than chance.) Greenleaf (24) when testing 471 children in Newton, found that the Newton Reading Readiness Test was too easy as a true measuring instrument of children's abilities and that the visual perception test had the greatest relation to reading achievement. It was very interesting to note in O'Sheasy's (39) study that teacher's judgements had the best prognostic value with the highest correlation of .65. The other measuring instruments used were the Gates Reading Readiness Test and the Pintner Cunningham Primary Test. Two-hundred-nine pupils participated in this study. Kent (31) made a study of three communities from the greater Boston area representing high, middle, and low socio-economic groups. He gave Thurston's Test of Primary Mental Abilities and the Metropolitan Achievement Test for Reading to 229 children and discovered that the Primary Mental Abilities Test is limited in its prognostic value and that there was a statistically significant difference between results for the high socio-economic group and middle group, but not between the middle group and low group. De Loura (14) testing 267 children of first grade and using the Pintner Cunningham Primary Mental Test and the Marion Monroe Aptitude Test as measuring instruments, reported that the I.Q. had the most important relation to readiness but not to reading achievement and mental age was limited

in its predictive value. Farrington (2) has made a study of the prognostic value of readiness and intelligence tests in seven first grades. Information concerning this may be found in the Readiness Chapter.

B. Relationship of Reading to Mental Tests.

Reading and its relationship to mental tests, especially among mentally retarded children, has been the topic of investigation in many studies. Haggerty (25) compared such tests of mental capacity as the Stanford Revision of the Binet-Simon Tests, Kuhlman Anderson, Henmon-Nelson and Durrell Sullivan Reading Capacity Tests with reading achievement measured by the Gates Silent Reading Tests on 233 fourth grade children, and concluded that in general there was a marked degree of relationship between reading capacity and achievement except in the case of the Stanford-Binet where it was low. Gilman (22) studied the mental test variations among dull children and analyzed the frequency of reading disability among the group using the Durrell Sullivan Hearing Capacity Test and the Stanford-Binet individual examination as criterion of ability each in comparison with the Durrell Sullivan Reading Achievement Test. The California Test of Mental Maturity and Pintner-Durost General Ability Tests were also used as measuring instruments. She obtained complete data on 189 special class children and found that there was a marked variation in the degree in which the tests agreed with each other, also that statistical evidence showed that the Binet mental ages and Hearing Capacity as criteria of reading ability fall short of their estimated capacity. Dodge (16) in a comparison of the Stanford-Binet mental ages and the

A

Hearing Comprehension measured by the Durrell Sullivan Hearing Capacity Test of 558 children in three metropolitan cities (Lynn, Mass., New Haven, Conn., and Rochester, N.Y.) reported that the difference between mental age and hearing comprehension was statistically significant and that dull children work up to and beyond their capacity if it is measured by mental age, but fall short of it if it is measured by hearing comprehension. Foss (20) in her study used a paired technique in comparing the various language skills of mentally retarded children and children with a foreign language background. She used the Stanford-Binet, Durrell Sullivan Reading Capacity, and Durrell Sullivan Reading Achievement Tests as measuring instruments. She discovered that there was a statistical difference in the scores in favor of the English speaking group in the hearing vocabulary and comprehension tests, but there was no significant difference on the achievement test. Cavanagh (9) made a comparison of the data from the following mental tests: (1) Otis Quick Scoring Mental Abilities Tests, Form A., Grades 1-4 and for Grades 4-9, (2) The Henmon-Nelson Tests of Mental Ability for Grades 3-8 and 7-12, (3) The Kuhlman-Anderson Tests for Grades 7-8 and 9 to maturity, (4) Otis Quick Scoring Mental Ability Tests, Gamma Test, Form C, Grades 9 through college. The tests were given to Grades, 4-8 and 9. Complete data was obtained on 140 cases. He discovered that the wide range shown by one test does not necessarily maintain itself in the result of other tests when comparisons of ranges are made and that the great percentage overlap between the intervals of the tests which indicate, that while the individual tests do give a posi-

tion of relative value they do not give an exacting discriminative value. Spencer (42) in her study analyzed the effect of reading disability on performance on the Terman Merrill Revision of the Stanford-Binet Intelligence Tests on 302 cases and found that scores on tests were lowered because of reading disability, but that the new revision lowered the amount of required reading over the first scale.

C. Analysis of Reading and Vocabulary Tests

A few studies have been devoted to analyzing established or standardized tests. Cavin (10) made a penetrating study of 20 intermediate grade reading tests and found great variations among them. She analyzed the tests according to: (1) abilities they measure, (2) nature of test items, (3) standardization of tests, (4) reliability of tests, (5) extent to which the tests discriminate between grades. After carefully studying the manuals accompanying each test, she reported that the stated abilities of the 20 tests measured differ widely in respect to number and kind and that there was insufficient information published in the manuals in respect to type of norms and reliability coefficients. She also discovered that the tests differed greatly in respect to their power of discriminating grades. The tests analyzed in the study were the following: (1) Detroit Reading Test, Test 111. Grades 4-6, (2) Durrell Sullivan Reading Achievement, Grades 3-6, (3) Gates Reading Survey, Grades 3-10, (4) Gates Silent Reading Tests, Grade 3-8, (5) Haggerty Reading Examination, Grades 6-12, (6) Ingram Clark-Diagnostic Reading Interest, Grades 4-8, (7) Iowa Silent Reading Test, Grades 4-9, (8) Los Angeles Elementary Reading Test, Grades 3-4, (9) Metropolitan Intermediate Reading Test, Grades 4-6, (10) Monroe

Standardized Silent Reading Test 1, Grades 3-5, (11) Nelson Reading Test, Grades 3-9, (12) New Stanford Reading Test, Grades 2-9, (13) Pressey Diagnostic Reading Tests, Grades 3-9, (14) Progressive Reading Tests, Elementary, Grades 3-6, (15) Public School Achievement Tests, Reading, Grades 3-8, (16) Sangren Woody Reading Test, Grades 4-8, (17) Shark Tests of Reading Scale, Grades 2-12, (18) Unit Scale of Attainment, Division 1, Grades 3-4, (19) Unit Scales of Attainment, Division 2. McAuliffe (34) made an item analysis of the Hurlburt Vocabulary Tests by studying 200 tests, 100 from each of grades nine and eleven and analyzing fifty of the top scores and fifty of the lowest scores. He discovered that adjective meaning had the lowest correct response and the poorest grade discrimination value. Many items also had poor validity. The other two parts of the test (nouns and verbs) had a better grade discrimination value and more valid items.

D. Comparison of the Vocabulary of Various Reading Series with Standardized Tests

Several studies have been made to determine the correlation between the vocabulary of different basal reading series with standardized tests. Fox (21) made a study of the percentage of words included in ten standardized reading tests that were also included in the pre-primer, primer, and first and second readers of the Scott Foresman Curriculum Foundation Series. The ten tests used in the study were 1) American School Achievement Tests, 2) Detroit Word Recognition Test, 3) Detroit Reading Test, 4) De Vault Primary Reading Test, 5) Gates Primary Reading Tests, 6) Lee Clark Reading Test, 7) Los Angeles Primary Word Recognition Test, 8) Metropolitan Achievement Tests, 9) Reilly Primary Reading Test, 10) Unit Scales of Attainment.

ACKNOWLEDGEMENT

Grateful acknowledgement is made for the help and guidance of our advisor, Dr. Donald D. Durrell.

First Reader: Donald D. Durrell

Second Reader: Helen A. Murphy

This study is part of a summary of research in the reading and language arts fields in which the experimental studies done at Boston University between 1930 and 1953 were reviewed. The complete study was done by: Evelyn Crocker, Ruth Cullen, John Deasy, Marjorie Follansbee, Sylvia Gavel, Claire Grossman, Mary Holmes, Ruth Lurie, Anne Melker, Viola Petrocelli, and Harriett Wilcox.

TEST CONSTRUCTION AND EVALUATION

Many studies at Boston University have been made in this area; some were concerned with evaluation of standardized tests and others have constructed their own tests and evaluated the material on a selected population. Although the classification of the different types of studies is difficult because it often overlaps, it is possible to divide the subjects into two broad areas: 1) those studies using established tests and 2) those studies that have constructed and evaluated new tests. The different areas that could be grouped under the established tests are: a) Evaluation of Tests for Prediction of First Grade Reading Success, b) Relationship of Reading to Mental Tests, c) Analysis of Reading and Vocabulary Tests, d) Comparison of the Vocabulary of Various Reading Series with Standardized Tests. Under group 2 the construction of new tests is a) Construction and Evaluation of Visual and Auditory Tests in the Primary Grades, b) Construction and Evaluation of Visual and Auditory Tests in the Middle Grades, c) Evaluation of Reading Tests, d) Construction and Evaluation of Vocabulary Tests, e) Construction of a Learning Rate Test, f) Construction of Spelling Tests.

I. Studies Using Established Tests

A. Evaluation of Tests for Prediction of First Grade Reading Success

The prognostic value of reading and intelligence tests for

After checking the word lists she discovered that the percent of words in both tests and the reading tests varied greatly. Macgregor (33) made a similar study using the same standardized tests and the Reading Foundation Series. After analyzing and comparing the vocabularies of the tests and books, she found that the tests varied in the number of words tested and in the sampling of the vocabulary also that no test had an adequate sampling of vocabulary for the Reading Foundation Series. Nason (36) using the same standardized tests compared their vocabulary with the Learning to Read Series. She also found great variation in number and percent of words among the vocabulary tests and that none of them was suitable as an accurate measure of reading achievement. Hughes (27) using the American School Achievement and the Metropolitan Achievement Tests studied their vocabulary in relation to three basal reading systems: 1.) Today's Work Play Books, 2.) Reading Foundation Series, 3.) Curriculum Series. She found that the percent of words included in the tests and the basal reading systems was not high, but that the Grade One words sampled by the tests was quite evenly divided among the three basal reading systems. Doherty (17) using the same ten tests that were used in the other study compared their vocabulary with the Ginn basal readers and found that none of the tests seemed to include a vocabulary which was representative of the Ginn basal readers. Jack (28) using the same materials came to the same conclusion. Cogan (12) compared the vocabularies of the Grade One and Two books of the "Alice and Jerry" series and "Today Work Play Book" series and studied their vocabulary in relation to the Gates Primary Reading test.

She found that the tests varied widely in the number and percent of words used. Young (46) using the same ten standardized tests that were used in previous studies compared their vocabulary with that of the Primary Social Studies books of the Curriculum Foundation Series. She discovered that there was a wide difference between the tests in percent of words common to the tests and books. De Wolfe (15) made a comparison of the ten standardized reading tests used in previous studies with the Easy Growth in Reading series and concluded that not one of the tests includes enough of the vocabulary of the basal reading series. Kennedy (30) using the same measuring criterion as was used in previous studies, compared the tests with the Reading for Interest Series and discovered that the number of words in the basal system which are tested is generally low. In all of the above studies the findings were similar, none of them found a high relation between the vocabularies used in the tests and the basal reading systems.

Construction and Evaluation of New Tests

A. Construction and Evaluation of Visual and Auditory Tests in the Primary Grades:

There have been several tests that have been constructed for the primary grades. These are concerned primarily with measuring auditory and visual discrimination. Kiernan (32) constructed a group auditory test of word analysis and checked its validity against tests of reading achievement and intelligence. Ninety-six second grade children participated in the study. After correlating the scores on the experimental test with scores on the Otis Mental Ability and the Stanford Primary Reading Tests, she found that the test constructed was too easy with the exception of the Phonograms Test which had a fair range and a good distribution of scores. The test showed some relation to reading achievement (the critical ratios of 3.80 and 6.53 were statistically significant) but little relation to intelligence, (none of the critical ratios were statistically significant.) Goggin (23) and others constructed group auditory tests for grades one, two and three and evaluated them by testing 234 first grade children, 217 second grade children and 187 third grade children. After statistically analyzing the tests they found that all the tests were good measures for grade one, fair for grade two and poor for grade three. After making an item analysis of the tests they discovered that there is a general increase in the knowledge of the sound of the letter with the increase in grade level, that the final sound of the letter was consistently the most difficult and that certain letters continued to present difficulty throughout the three grades (l-n-r.) Group tests in reading were

constructed by Bennet (2) and others and were evaluated by testing 826 children in grade one, 787 in grade two, and 766 in grade three. After making an item analysis of each of the items of the test they discovered that most of the items in all four tests showed a statistically significant critical ratio. One Hundred Thirty out of 803 items were not statistically significant. Various measures to analyze the knowledge of letter names in grades one, two and three were developed by Casanove (11) and others. Five group tests and one individual test was given to 607 children in grades one, two and three in 21 classrooms in 17 different communities representing low, middle and upper socio-economic communities. When scores were analyzed and correlated they discovered that the children in grades two and three had a good knowledge of letter names and that the tests showed a wide range in grade one and fairly high reliability. Jenkins (29) compiled a series of exercises to develop visual discrimination in beginning reading and taught them to an experimental group of 49 children for a period of six weeks. In order to evaluate them she developed tests to measure visual discrimination, auditory discrimination, and learning rate of the experimental group and the 46 children in the control group. When she analyzed and compared the scores she discovered that in the visual discrimination and learning rate tests the experimental group was superior, but in the auditory test the control group showed superior gains.

B. The Construction and Evaluation of Visual and Auditory Tests in the Middle Grades

Hogan (4) has developed visual and auditory discrimination

tests for grades four, five and six. Information concerning this may be found in the chapter on word analysis.

C. Evaluation of Reading Tests

Daniels (13) has made a study of informal reading tests built on three basic reading series, ("The Child Development," "Alice and Jerry," and "The Children's Own Readers" series.) Standardized tests such as the Metropolitan Achievement, Gates Primary, Gray Oral Reading Paragraphs, Durrell Oral Reading Paragraphs, Durrell Sullivan Reading Capacity and the Gates Silent Reading Test were used on 246 children in grades one, two and three. After the scores were analyzed and inter-correlations were made between the ratings given by the room teachers, Supervisor and assistant with the composite score on several tests, she found that the assistant showed the lowest correlations and the classroom teachers showed the highest correlations between their ratings and the composite score. Waters (44) analyzed a Readiness Battery Pre-Reading Test of Ginn and Co. and evaluated it by giving the tests to 656 first grade children of twenty-five classes from eleven different states. After analyzing the scores and computing correlations between the experimental test and such intelligence tests as the Stanford Binet, Kuhlman Anderson, and Detroit First Grade Intelligence, she discovered that there was a skewed distribution of scores in the test and that there were low correlations between the test and the intelligence tests, indicating that the experimental test measured different factors. McCarthy (6) constructed and evaluated a Reading Readiness test. For additional information about this refer to the chapter on Readiness.

D. Construction and Evaluation of Vocabulary Tests

The construction and evaluation of vocabulary tests and exercises to increase children's use of descriptive words has been the source of interest in many of the studies. Simon (41) by using both a multiple response and a matching technique in testing 200 children, tried to measure various meanings and shades of meanings, plus synonyms. When the item analysis was made she noted that the test had many weak items, although it was able to help in discrimination between the poor and able readers. Nott (38) also was interested in measuring a variety of word meanings and in addition tried to do it in a shorter space and time. Three hundred children participated in the study, 100 from grades 5,6 and 7 respectively. The scores of the test were correlated with the scores of the Metropolitan Achievement to see what relationship, if any, existed between the tests. After computing the correlations Nott reported that the test established a successful upper limit and that it proved to be too difficult for grade five. Wilking (45) was concerned with constructing a vocabulary test that could be incorporated in a diagnostic reading test. He used the category method in testing and received his items from reading books by Durrell and Sullivan and the first and twentieth thousand words in the Teacher's Word Book. After building the experimental test it was evaluated on 356 students in grades three through thirteen. The reliability was tested by a multiple correlation technique and the validity was checked by another criterion test. After studying the test results he reported that the test was able to distinguish between grades and that it was a highly valid measure

of reading vocabulary as the reliability coefficient .69 indicated. Euphrosyne and MacPartlin (18) conducted an experiment with 108 third grade children to find out whether or not a period of six weeks teaching using constructed exercises would improve and increase a child's use of descriptive words. After six weeks teaching with the experimental group the control and experimental groups were tested to see if there was any statistical difference between groups. They found that in regard to the use of descriptive words the experimental group had a superior score, but in non-descriptive words the critical ratio of 5.18 was in favor of the control group.

Bertoli (3) constructed and evaluated a dictionary test for use in Grades four, five and six. After administering the tests to 176 children in those grades she found that the test was too easy for grades five and six and would be more suitable for grade four. Varney (43) constructed five tests of recognition type using matching multiple choice and the true false techniques, each testing various meanings of 20 words. The tests were given to 155 children in Grades four, five and six in a large residential community. After correcting the tests she discovered that the matching and multiple choice techniques were valid and the true false technique had the lowest item validity. McNiff (35) constructed two equivalent forms of an intermediate grade vocabulary test to measure word recognition and meaning stressing a wide sampling of vocabulary and maintaining an economy of time and space. The tests were administered to 431 children in grades four, five and six

in three suburban communities of varying socio-economic backgrounds and she found that the tests met the requirements of the problem. Berwick (5) developed and evaluated multi-meaning vocabulary tests for grades four, through eight. She gave pre-liminary tests to 500 pupils in grades four through eight revised them and compiled new forms. After scoring the tests and making an item analysis she found that the tests were highly valid and reliable.

E. The Construction of a Learning Rate Test.

Several studies, dealing with learning rate, have been made and one of these is a test construction and evaluation. Farr (19) constructed two tests, Identification Check and Picture Check to be used in Grade one after a controlled teaching period. She correlated these scores with the Metropolitan Readiness, Gates Reading and Metropolitan Achievement tests to see what relationship, if any, existed between learning rate and those factors. After correlating scores she found that the correlations between learning rate and the readiness and achievement tests was low and that there was no relative difference between the Picture Check and Identification Check.

F. The Construction of Spelling Tests

Spelling, and its relation to reading, has long been a source of inquiry at Boston University and many tests have been constructed in this field. Carter (8) constructed tests of auditory discrimination, visual perception, and speed of handwriting for the middle grades in order to diagnose inadequate performance in spelling and to see what relationship, if any, existed between the factors tested and spelling achievement. In addition to the constructed tests she gave 180 pupils in grades five and six the Stanford Achievement Test of Spelling, the Metropolitan Achievement Test of Spelling and the Kuhlmann Anderson Test of Intelligence. She then correlated the scores with the spelling achievement, made an item analysis and drew up scatter diagrams to determine the importance of the tested factors to reading achievement. The results of the data indicated that auditory discrimination and visual perception had statistically significant correlations with the spelling achievement, while speed of handwriting and mental age did not; the item analysis showed that the test items had the proper range of difficulty and the scatter diagram indicated that auditory discrimination and visual perception are important factors in spelling achievement, although the relationship is not perfect. Biggy (6) constructed a modified sentence type of spelling test and a recognition type of spelling test and compared them to determine which yielded the best index of spelling difficulty. In addition the Niles Visual Test and Auditory Tests were given and correlated with the constructed tests. These four tests were given to 2,645 children in the fifth grade. After the

test scores were correlated and the reliability of the constructed tests was established, it was found that approximately sixteen percent of the population correctly spelled just slightly better than half of the words presented and that both the modified-sentence test and the recognition test was highly reliable as indicated by the correlations .978 and .98. She discovered that the correlation between the two methods of measuring spelling ability was .6, comparatively high, but not high enough to suggest that the two methods are interchangeable. Nichols (37) constructed a group test for the purpose of detecting spelling difficulties and locating their cause. She constructed two forms, Form A was given to 1,615 pupils in grades three, four, five and six, Form B was given one month later to 1,524 children. The two forms contain the following subtests. I, Spelling Achievement, II, Proof Reading, III, Word Meaning, IV, Handwriting, V, Visual Discrimination, VI, Auditory Discrimination. After the scores were correlated and validity tested she found that the test was valid and reliable and that the various sub-tests showed fair correlation with spelling achievement, except in handwriting. Sherman (40) constructed and analyzed a spelling readiness test, to determine when a child is able to begin formal spelling. The test consisted of the following items, I. Letter Form, II. Initial Consonants, III. Final Consonants, IV. Initial Blends, V. Final Blends, VI. Phonograms, VII. Visual Memory. The test was given to 141 pupils of grade one in five classes and four cities and suburban towns. After the scores of sub-tests were correlated and the validity checked, she found that out of 97 test items 86 had a critical

ratio of 3.000 or above, and that there was not significant differences between the boys and girls on any section of the test. Barrett (1) and others constructed tests of Auditory Discrimination, Visual Discrimination, Visual Analysis, Writing from Visual Memory, Writing from Auditory Stimulus, Handwriting and Auditory Analysis to determine what relationship these factors had to spelling achievement. The tests were given to 468 pupils in the fourth, fifth and sixth grades in five schools and the results were correlated with the spelling scores from the Metropolitan Achievement Tests. The pupils in each grade were paired for sex, M. A., C. A. and reading grade with at least a years spelling difference. Twenty pairs were selected who showed a constant C. A., M. A. and reading grade, but a significant difference in spelling above and below grade. When they analyzed the data they found, that auditory and visual discrimination bear a high relation to spelling ability and auditory analysis and speed of handwriting showed the least relationship. Brooke (7) constructed tests of recognition and recall in measuring visual and auditory perception and correlated them with the scores on the Stanford Achievement Test to study their relationship. After testing 180 children in Grades four, five and six and correlating the scores she found that the recall was superior to the recognition test. Conway (1) and others constructed tests of visual perception, auditory perception, and a phonetics test for Grades two and three. Hill (3) constructed visual discrimination, auditory discrimination, and motor coordination tests. For additional information concerning these studies refer to the chapter on Spelling.

Bibliography

1. ✓ Barrett, Patricia et al. The Relation of Perceptual Factors And Speed of Handwriting To Spelling Ability. Ed. M.
2. ✓ Bennett, Virginia et al. Construction And Evaluation of Group Tests in Reading For Grades One, Two and Three.
3. Bertoli, Lillian. The Construction and Evaluation of A Dictionary Test. Ed. M. 1949. 49p.
4. ✓ Berwick, Mildred M. An Evaluation of the Prognostic Values of Certain Pre-Reading Tests To Reading Achievement. Ed. M. 1947. 53p.
5. ✓ Berwick, Mildred M. The Construction And Evaluation Of Multi-Meaning Vocabulary Tests For Grades Four Through Eight. Ed. D. 1952. 224p.
6. ✓ Biggy, Mary Virginia. A Comparison of The Recognition Types of Spelling Test With The Modified Sentence Type As A Means Of Determining The Best Index Of Spelling Difficulty Of A Selected List of Words. Ed. D. 1953. 251p.
7. ✓ Brooke, Barbara Ann. Comparison Of Recognition And Recall Of Word Elements As Methods Of Measuring Visual And Auditory Perception In Spelling. Ed. M. 1947. 68p.
8. Carter, Bernadette. The Construction Tests of Visual Perception, Auditory Discrimination and Kinesthetic Factors To Be Used In Diagnosing Inadequate Performance In Spelling. Ed. M. 1941. 71p.
9. Cavanagh, Francis James. A Comparison of Data From Selected Mental Tests. Ed. M. 1944. 57p.
10. Cavin, Grace E. An Analysis of Intermediate Grade Reading Tests. Ed. M. 1940. 78p.
11. ✓ Cazanove, Frances et al. The Construction And Analysis Of Tests To Measure The Knowledge Of Letter Names In Grades One, Two and Three. Ed. M. 1953. 38p.
12. ✓ Cogan, Anna M. A Comparison of The Vocabularies Of The Grades One And Two Books Of The Alice And Jerry Series, The Curriculum Foundation Series, The Today Work Plan Book Series And The Gates Primary Reading Standardized Test. Ed. M. 1950. 55p.

13. Daniels, Katherine H. An Evaluation Of Certain Informal Reading Tests. Ed. M. 1940. 70p.
14. ✓ De Loura, Bertha Amy. A Study Of The Predictive Value Of Intelligence And Aptitude Tests To Reading Achievement In Grades One. Ed. M. 1949. 71p.
15. ✓ De Wolfe, Barbara Ellen. A Comparison Of The Vocabularies Of The Easy Growth In Reading Series And Ten Standard Reading Tests. Ed. M. 1950. 55p.
16. Dodge, Harriet Eva. A Comparison Of The Stanford-Binet Mental Age And The Hearing Comprehension Ability Of Dull Children. Ed. M. 1940. 33p.
17. ✓ Doherty, Mary Louise. Comparison Of Ten Standardized Tests With The New Work Play Books. Ed. M. 1948. 81p.
18. ✓ Euphrosyne, Georges, MacPertlin, Anna. Evaluation Of Exercises Designed To Increase Children's Use Of Descriptive Words. Ed. M.
19. Farr, Carol V. Construction And Avaluation Of Tests Of Rate Of Learning In Reading For Children In First Grade. Ed. M. 1942. 55p.
20. Foss, Gertrude Mae. Language Comprehension Skills Of Mentally Retarded Children. Ed. M. 1938. 37p.
21. ✓ Fox, Ruth E. A Comparison Of Ten Standardized Reading Tests With The Curriculum Foundation Series. Ed. M. 1948. 52p.
22. Gilman, Alice. Mental Test Variations Among Dull Children. Ed. M. 1941. 104p.
23. ✓ Goggin, Dorothea D. et al. Construction And Statistical Analysis Of Group Auditory Tests For Grades One, Two And Three. Ed. M. 1953. 89p.
24. Greenleaf, Edith Eleanor. An Evaluation Of Visual Perception Tests For Predicting Success In First Grade Reading. Ed. M. 1936. 34p.
25. Haggerty, Earl J. An Evaluation Of Certain Tests Used As Measures Of Reading Capacity. Ed. M. 1940. 74p.
26. ✓ Haskell, Barbara A. et. al. The Relationship Of The Knowledge Of Letter Names And Reading Achievement In Grade One. Ed. M. 1952. 31p.
27. ✓ Hughes, Alice A. An Analysis Of The Vocabulary Of Two Standardized Reading Tests In Relation To The Vocabulary Of Three Reading Systems. Ed. M. 1950. 61p.

28. ✓ Jack, Barbara A. A Comparison Of The Vocabularies Of Ten Standard Achievement Tests In Reading With The Vocabulary Of The Ginn Basic Readers. Ed. M. 1948. 63p.
29. Junkins, Kathryn M. The Construction And Evaluation Of Exercises For Developing Visual Discrimination In Beginning Reading. B. S. in Education, State Teacher's College, Winona, Minn., 1939.
30. ✓ Kennedy, Helen Ursula. An Evaluation Of Ten Standard Achievement Tests In Reading In Comparison With The D. C. Heath Basic Readers. Ed. M. 1948. 44p.
31. ✓ Kent, Helen H. An Evaluation Of A Test For Predicting Reading Success In Grade 1. Ed. M. 1948. 66p.
32. ✓ Kiernan, Anna Constance. The Construction And Evaluation Of An Auditory Test Of Word Analysis. Ed. M. 1950. 47p.
33. ✓ Macgregor, Helen Denison. An Analysis Of The Vocabulary Of Ten Selected Reading Tests And The Vocabulary Of The Reading Foundation Series. "The Alice And Jerry Books." Ed. M. 1948. 120p.
34. ✓ McAuliffe, Anita Serena. An Item Analysis Of Hurlbyrt Vocabulary Tests. Ed. M. 1948. 168p.
35. ✓ McNiff, Margaret Ellen et al. The Construction Of A Vocabulary Test For The Intermediate Grades. Ed. M. 1953. 95p.
36. ✓ Nason, Doris E. A Comparison Of The Vocabularies Of The Grade One And Two Books Of The Learning To Read Series And Ten Standardized Reading Tests. Ed. M. 1948. 88p.
37. ✓ Nichols, Augusta Matilda. The Construction And Use Of A Group Test For The Analysis Of Spelling Difficulties. Ed. M. 1947. 220p.
38. Nott, Helen Sullivan. The Construction And Evaluation Of A Reading Vocabulary Test For Grades V, VI and VII. Ed. M. 1950. 59p.
39. O'Sheasy, Edward Andrew. A Study Of The Predictive Values Of Certain Pre-Reading Tests Employed At The First Grade Level. Ed. M. 1951. 76p.
40. ✓ Sherman, Dorothy E. The Construction And Analysis Of A Spelling Readiness Test. Ed. M. 1950. 52p.
41. Simon, Margaret. The Construction And Evaluation Of A Reading Vocabulary Test For The Middle Grades. Ed. M. 1951. 51p.

42. Spencer, Doris. The Effect Of Reading Disability On Performance On The Terman-Merrill Revision Of The Stanford Binet Intelligence Tests. Ed. M. 1943. 54p.
43. Varney, Elinor Caroline. An Experiment To Evaluate Some Techniques For Measuring The Knowledge Of Word Meanings In The Fourth, Fifth, And Sixth Grades. Ed.M. 1945. 84p.
44. ✓ Waters, Mildred. A Partial Evaluation Of An Experimental Pre-Reading Test. Ed. M. 1949. 139p.
45. Wilking, Stephen Vincent. The Construction And Evaluation Of A Measure Of Reading Vocabulary. Ed. M. 1940.
46. ✓ Young, Patience. A Comparison Of The Vocabularies Of Ten Standard Achievement Tests In Reading With The Primary Social Studies Books Of The Curriculum Foundation Series. Ed. M. 1951. 119p.

Gross Reference

1. Conway, Virginia et. al. A Study Of Factors Related To Primary Grade Spelling And Correlations Of Factors Studied. Ed. M. 1951. 135p.
2. Sargington, Julia H. The Prediction and Evaluation of Readiness and Intelligence Tests in Seven First Grades. Ed. M. 1949. 56p.
3. Hill, Barbara. The Persistence Of Perceptual Functions Related To Reading And Spelling. Ed. M. 1943. 97p.
4. ~~Hagan~~ Hagan, Irene Marie, The Construction of Tests in Visual and Auditory Discrimination for Grade Four, Five and Six. Ed. M. 1952. 79p.
5. Johnson, Mary Elizabeth. Study Of The Understanding Of Vocabulary Of Content Subjects By Children Of Grade Five. Ed. M. 1950. 91p.
6. McCarthy, Josephine. The Construction and Evaluation of a Test of Reading Readiness. Ed. M. 1941. 42p.

*

Barrett, Patricia et. al. The Relation of Perceptual Factors and Speed of Handwriting to Spelling Ability. Ed. M. 1951. 77p.

Purpose: To determine the possible relationship between spelling achievement and visual identification of words flashed, visual analysis of word elements, writing from visual memory, auditory word identification, auditory analysis of words, writing from auditory stimulus and speed of handwriting from copy.

Materials: A battery of seven tests were given to test the above factors. Sample of the tests are included. The tests were given to a group of 468 pupils in the fourth, fifth, and sixth grades in five schools.

Tests used in Barrett Study of Spelling Ability.

Visual Identification of Words Flashed.

It consisted of 30 words from easy to difficult. A word was flashed for five seconds. Then the word was removed, the child was told to wait for five seconds then to write the word.

Samples of the words used:

- | | |
|---------------|-----------------|
| 1. knelt | 19. intrinsic |
| 6. lintel | 25. pronounce |
| 15. ferocity. | 30. geometrical |

Visual Analysis of Word Elements

A word is flashed five seconds and after five seconds the child crosses out all letters he can remember seeing. All parts

of the word can be found but not in the proper order.

Sample:

1. key word: complement

a ment com ab n ion ple

21. key word: disagreeable

a a b gree ble dis age

Writing from Visual Memory

There were 36 exercises with from five to seven words in each exercise. The words are shown from five to eight seconds, and then the children were to cross out the word they had seen.

Sample:

1. key word: dog

ago dog do go girl

9. key word: weather

other then wealthier weather whether wither
wealth

30. key word: accordance

ascendance accordion accomodate
accidence accordance accusative

Auditory Word Identification

In this test the examiner speaks a word, the child finds that word on his paper and draws a line through it. The child always had a choice from five words.

Sample:

1. key word: boy

dog day boy boys play

32. key word: insolvent

insolent incessant insolvent solvent insurgent

Auditory Analysis of Words

The examiner speaks a word and the child then finds on the paper all the sounds he has heard and circles them. Five seconds is allowed for circling.

Sample:

1. key word: blinding

ing r bl sh t ing

10. key word: voyage

oy re j v l t

Writing from Auditory Stimulus

The child writes the word the examiner says. Most of the words were new to the children. Phonetic spelling was acceptable. They were told to spell them the way they thought the words sounded. Credit was given for each syllable correctly or phonetically spelled.

Speed of Handwriting:

The child was asked to copy a paragraph. He was stopped after two minutes and his letter speed per minute was then found.

Procedure: The tests were given and the results were correlated with the spelling scores for the same children from the Metropolitan Achievement Tests. The pupils in each grade were paired for sex, M.A., C.A., and reading grade with at least a year's spelling difference. Both could be below or above grade. There were 82 pairs found. The 82 pairs were sifted to retain only those who showed a constant C.A., M.A., and reading grade but a significant difference in spelling above and below grade. There were 20 such pairs.

Analysis of Data: The tests correlated with spelling ability as follows:

1. Auditory Discrimination	.708
2. Visual Discrimination	.703
3. Visual Analysis	.654
4. Writing from Visual Memory	.626
5. Writing from Auditory Stimulus	.621
6. Handwriting	.360
7. Auditory Analysis	.350

In the matched pairs despite controlling for reading the results showed a marked difference in spelling ability with visual identification showing the most difference: a mean difference of 10.7 with the 82 pairs and a mean difference of 16.2 with the 20 pairs.

Conclusions: Auditory and visual discrimination bear a high relation to spelling ability, this is consistently so with visual discrimination. Auditory analysis of word and speed of handwriting showed the least relationship.

*Barrett, Patricia

Lawton, Helen Randall

Bennett, Priscilla Ann

Murphy, Bartholomew H.

Flemming, Catherine

Newton, Anne

Kyres, Thalia

Bennet, Virginia E., et. al.* Construction and Evaluation of
Group Tests in Reading for Grades One, Two and Three. Ed.
M., 1953. 297p.

Problem: To construct and evaluate a group test in reading for grades 1-2 and 3, which will include a large vocabulary from primary basal reading series and measure the various skills taught in the reading program.

Procedure:

(1) The vocabulary from six widely used basal systems was assembled and compiled to make a common vocabulary. The basal systems used were:

- (a) Scott Foresman and Company
- (b) Row Peterson and Company
- (c) Houghton Mifflin and Company
- (d) Ginn and Company
- (e) Macmillan and Company
- (f) Silver Burdett and Company

(2) The exercises used in these systems were analyzed and they were found to have the following items in common:

- (a) Classification of words, phrases, sentences and paragraphs
- (b) Word recognition
- (c) Main idea
- (d) Sequence
- (e) Context clues
- (f) Association
- (g) Major and minor ideas

(3) It was decided to build items to measure these skills. The group divided into four pairs and each pair constructed a form for each of the three grades. Four forms called A-B-C-D for each grade were constructed making a total of twelve tests.

(4) All forms of tests were administered in each grade. Data was obtained from a total of 2,379 pupils, 826 in Grade 1, 787 in Grade 2 and 766 in Grade 3.

(5) An individual item analysis was made of each test at each grade level of the 50 highest and lowest scores. The percent of correct responses was found for each group. Edgerton table were used to find the standard error squared for each item. This was followed by the computation of the standard error of the difference and the critical ratio for each item.

Conclusion:

(1) The scores of the tests administered indicated that children can adequately handle items of this type in grades 1, 2 and 3.

(2) Most of the items in all four tests showed a critical ratio which was considered statistically significant. Out of the total number of items in all four tests (803) 130 of those 803 items were not statistically significant.

(3) In form A the difficulty of the section as going from easy to more difficult was classification, sequence, main idea, major and minor idea, and context clues.

(4) The difficulty of the sections going from easy to more difficult in form B is as follows: word recognition, association (Grades 2 and 3) classification, major and minor ideas, context clues and word meaning. The responses seem to indicate that these skills are more difficult for Grade 1 than Grades 2 and 3 with the exception of word recognition and classification of words.

(5) In form C of the 2nd, grade test, phrases should be added to the classification and association tests to increase the difficulty. There is not a wide enough difference in percent of responses between the high and low pupils.

*

Elsie M. Coffey
Joan R. Cooney
Elaine K. Cotter
Anngenetie Groton
Mary E. Howard
H. Jeanette O'Brien
Anna M. Palmer
Marguerite Belleau

Bertoli, Lillian. The Construction and Evaluation of a Dictionary Test. Ed.M. 1949. 49p.

Problem: To construct and evaluate a dictionary test suitable for use in Grades IV, V, and VI.

Procedure: 1. A list of basic skills was drawn by the author and six tests were constructed to measure these skills. They are as follows: (a) Test I, Part A - consists of a box of words that are to be arranged in alphabetical order. Box 1 consists of different initial consonants; Box 2 consists of similar initial consonants and Box 3 consists of similar consonant blends. In the fourth box the first three consonants of the words are similar. Test I, Part B - Groups of words are given and it is necessary to detect the one word in each group which is out of order. (b) Test II: Pairs of guide words are put on the paper. Two words are under each guide word. The two words that do not belong under the guide word are to be crossed out. (c) Test III: Definitions: A series of sentences is provided with one word underlined in each sentence. A sample dictionary is included from which the best meaning for each underlined word is to be chosen. (d) Test IV: Dividing Words: A series of words divided into syllables is given. The correct syllabication is to be given. Test IV, Part B - A list of words is given and the number of syllables in each word is to be written next to it. (e) Test V: Accent Marks: Words with a series of numbers written after them are listed. The number of the syllable that is accented is encircled:

e.g., company (1) 2 3 4 5 6. (f) Test VI: Pronunciation: A list of words is presented which is not in the speaking vocabulary of the children. The children have to use the diacritical marks in order to determine the correct pronunciation of the word and then mark the rhyming word, e.g. - a pri. cot (a pri-kot; ap ri-kot) apricot coat tot rete. The children have to decide how to pronounce apricot and then find the corresponding word that rhymes with it.

2. The test was given in December, 1948 to 186 children in Grades V and VI. The test was considered too difficult to be given to Grade IV.

3. Charts were drawn up to see the frequency distribution of scores for each test. Tables were drawn to show the differences between grades tested.

4. After charts and tables were drawn up an item analysis was made of Test 1, Parts A and B and Test 11 because they showed the greatest difference between grades tested.

Conclusions: 1. All the means for Grade VI were higher than the means for Grade V by a difference of 4.00 points in the first part of the test of alphabetical order to a difference of .053 on the pronunciation test.

2. All the means fell at a point that was above 50% of the possible score, except the means for Test 11. Guide Words (Grade VI 22.53 Grade V 19.44) The mean for Grade VI fell at a point of 49% below the possible score of 44 and for Grade V the mean fell 66% below the possible score, and Test VI Pronunciation in Grade V (6.90).

3. The high means in relation to possible score in Test 1-B (Alphabetical Order), Test 111 (Definitions), Test IV - A-B (Dividing Words) and Test V (Accent Marks) in Grade VI indicate a need for more difficult items at this level. Test 1 - M 9.26 in relation to possible score of 12. Test 111. M. 18.24 out of possible score of 24. Test IV. Part A. M. 9.97 83.1% of the total score of 12. Test IV. Part B., M. 6.71 out of a possible score of 8.54% received a perfect score. Test V. M. 18.43 out of a possible score of 23. 35% had a perfect score.

4. In Grade V the means were rather high for Test IV, (Part A. 9.24 - Part B 6.56) and Test V (17.04)

5. Test 11 (Guide Words) and Test VI (Pronunciation) were within the ability of both grades and difficult enough to provide a ceiling for discriminating the superior pupils of the group. In Test 11 there was only one perfect score in Grade VI with only three pupils in all making scores from 40 to 44. The highest score in Grade V was 39 with the next highest score being 33. The lowest score was 33. In Test VI the range of scores was 0-14 (14 highest possible score) in Grade VI and from 2-13 in Grade V. Only one pupil in Grade VI had a perfect score.

6. An item analysis of Test 1 and 11 resulted in a discovery that 23 items ought to be replaced by more difficult items. In Test 1 Part A items 1,2,3,4,5,6,11,12,13 and 22 were done correctly by 98-100% of the top pupils. In Part B items 1,2, 9 all were passed at a 100% by the top pupils. The first ten items in Test 11 should be replaced, especially items 5,7 and 9, which had no errors among the fifty highest pupils.

7. An analysis of the data shows that the test probably would have been better for Grade IV as originally planned.

Berwick, Mildred M. An Evaluation of the Prognostic Value of Certain Pre-Reading Tests to Reading Achievement. Ed.M.
1947. 53p.

Problem: To study the prognostic values of certain pre-reading tests in their prediction of reading efficiency.

Procedure: 1. In October, 1945, the Kuhlmann-Anderson Group Intelligence Test for Grade 1, Form B was given to 238 first grade children.

2. In January, 1946, the Lee-Clark Reading Readiness Test, 1943 Revision, was given to the same children.

3. In May, 1946 the Gates Primary Reading Test for Grades 1 and 2, Form 3, was given to determine reading achievement.

4. The following data was compiled for each child:

- a. Chronological Age
- b. Mental Age
- c. I.Q.
- d. Kuhlmann-Anderson Intelligence Test Score
- e. Lee-Clark Reading Readiness Test Score
- f. Gates Primary Reading Test Score
- g. Reading Grade and Reading Age

5. Percentage frequency distribution tables containing the mental ages, reading readiness test scores were set up. Means and Standard Deviation of the Means were computed for mental age, I.Q.'s, and the scores on the Lee-Clark Reading Readiness Test and the Gates Primary Reading Tests.

6. The predictive value of the tests were determined by correlating the scores made on the two pre-reading tests with the scores made on the reading achievement tests. Correlations were computed between the sexes and reading achievement and between

the different intelligence groups and reading achievement.

7. Comparisons were also made on the first two sections of the readiness test and the last two sections to determine which items gave the most reliable measurement.

Conclusions: 1. There was an indication that a knowledge of a child's mental age was an aid in predicting his probable success or failure in reading, as the correlation between mental age and reading achievement was .478 with a standard error of $\pm .050$. This indicates a predictive value of about 13% better than chance,

2. I.Q.'s had about the same predictive value as the factor of mental age. The correlation was $\pm .489$ with a standard error or $\pm .043$. The above correlation implies an efficiency of prediction of 13% better than chance.

3. The specific Reading Readiness Test used seemed to have as high a predictive value as the factors of mental age and intelligence quotients. The correlation of the readiness test with the achievement test was $\pm .466$ with a standard error of $\pm .050$. This indicates a forecasting efficiency of 13% better than chance.

4. There was a slight difference in the prognostic value of Subtests 1 and 11 compared with the values of Subtest 111 and 1V. Subtests 1 and 11 was about 10% better than chance. Subtests 111 and 1V was about 8% better than chance.

5. Girls and boys' scores had about the same predictive value. The boys' correlation was $\pm .479$ with a standard error of $\pm .071$. The girls' correlation was $\pm .450$ with a standard error of $\pm .072$. The predictive value of both was about 13%

better than chance.

6. The correlations found between reading achievement and the factors obtained from the results of these two pre-reading tests ranged from $r = .125$ to $r = .489$. The percentage of forecasting efficiency of these correlations was from five to 13 per cent better than chance which indicates that these predictive tests were limited in their use as prognostic instruments. It would seem that if the purpose in giving the tests is just to find out how many children will succeed or fail in first grade reading, it would not be necessary to give more than one test, since both tests yield the same forecasting efficiency (13% better than chance). It is possible, however, that the tests measure different skills and by giving them, more specific information might be obtained.

Berwick, Mildred Margaret. The Construction and Evaluation of Multi-Meaning Vocabulary Tests for Grades Four - Eight.
Ed. D. 1952. 224p.

Problem: To develop and evaluate multi-meaning vocabulary tests for grades four through eight to measure the number of common concepts individuals have for words that have more than one meaning.

Materials:

1. Three hundred twenty-one words from Thorndike's Word Book of 30,000 Words.
2. Webster's New International Dictionary of the English Language.
3. Thorndike-Lorge Semantic Count of English Words.
4. Walker-Cohen test of significance.

Procedure:

1. Preliminary testing was done on a sample population of 500 pupils in grades four through eight.
2. The preliminary tests were made with three sub-tests. The first was a recognition test of 803 words. The children showed their abilities to distinguish between single-meaning and multi-meaning words. The second was an identification test of 321 words and 1984 meanings. The children here picked out from a list containing some distractor items the common meanings for the multi-meaning words. The third test consisted of 96 paragraphs

and 294 meanings. Here the children tried to get meaning from context. Each test contained units that were rotated so that no items would be favored.

3. Scoring was done by ten teachers experienced in test correction.
4. Difficulty and discriminating levels were then established by means of an item analysis.
5. Tests were revised on the basis of the data thus obtained.
6. The final forms of the recognition test consisted of 484 multi-meaning words organized into three forms. The identification test included 151 words and measured 690 meanings in the three forms. The context test was made up of 61 paragraphs and tested 183 meanings in the three forms.
7. Final forms of the tests were not administered, but rescoring was done on each of the items accepted for the final tests from the initial tests. The rescoring consisted of these steps:
 - a. Making of new keys for the items accepted for the final forms of the test.
 - b. Rechecking each test and rescoring each item.
 - c. Giving each correct response a value of one.
 - d. Tabulating each pupil's total raw score on each form of the test.
 - e. Recording data for each pupil on individual cards which gave all information concerning his performances on all the tests given for this study.

Major Findings and Conclusions: A very thorough analysis of data of these types was made: population data, mean and standard deviation data, critical ratio data, validity data and reliability. The analysis indicates:

1. More words were tested in approximately the same time and space as a single meaning vocabulary test usually tests.
2. The tests were highly valid. The average coefficient of correlation between scores on the constructed tests and the mental age was .75 and between reading age and the total scores on the constructed tests .85.
3. The different forms and the sub-tests showed high reliability. The average coefficient of reliability between the scores obtained on Form A and Form B was .85; Forms A and C .84; Forms B and C .83. Between the sub-tests of the constructed test the coefficient of correlation was .86 between the recognition and identification test and .85 between the recognition and context and between the identification and context test.
4. A more substantial inventory of a pupil's knowledge of words which have more than one meaning was secured.

Biggy, Mary Virginia. A Comparison of the Recognition Type of Spelling Test with the Modified-Sentence Type as a Means of Determining the Best Index of Spelling Difficulty of a Selected List of Words. Ed. D., 1953. 25lp.

Problem: To compare the modified-sentence type of spelling test with the recognition type of spelling test to determine which yielded the best index of spelling difficulty. It was, therefore, necessary to construct recognition and modified sentence tests.

Materials: The manner of selecting words for the tests is discussed under procedure.

Definition of terms:

Modified-Sentence Test. Modified-sentence test is that type of recall test which requires the teacher to read the key word, read an illustrative sentence in which the word is used and then repeat the key word. The student writes only the key word.

Example: run John had to run to the store. run

Recognition Test. In this study, the recognition tests are built around 300 key words, all of which appear correctly at least once and incorrectly twice. The student is asked to identify the word as correctly or incorrectly spelled.

Visual Test. A visual test in the sense it is used in this experiment connotes the testing of visual perception by items designed to require careful visual discrimination on the part of the student in order to choose from like forms, the one that has been exposed to his vision for five seconds.

Auditory Test. In this study, the auditory test was used to determine the ability of each pupil to hear sounds in words.

Preliminary testing was done with three fifth grades apart from the population used in the experiment. For the testing to be analyzed, 2,645 cases were gathered. The following tests were given: five forms of the Biggy Modified-Sentence Spelling Test, the Biggy Spelling Recognition Test, the Niles Auditory Test, and the Niles Visual Test.

Procedure: Preparing the commonest list of words possible was done by examining ten commercial spelling series published between 1940 and 1949. From this list the commonest one were selected by use of eight well known spelling lists. A total of 326 words appearing in seven or eight of the lists was reduced to 300 by excluding 26 lacking social utility. These 300 words formed the bases for the tests.

Each word was given a six-place random number from the Acomb-Durost Table of Random numbers and then sorted in numerical order. The first sixty cards drawn according to this order constituted the first form of the Biggy Modified-Sentence Test. A similar process yielded five other forms of sixty words each.

The construction of the Biggy Spelling Recognition Test required the choice of two plausible misspellings for each of the 300 words. The same frame sentences were used in developing the items for the recognition test. A pattern calling for the presentation in one sentence of the two key words in the sentence both wrong, then both right, then one right and one wrong and then the reverse produced 600 sentences. The same random order system was used with the first 100 cards drawn constituting Form I and so on for the other five forms.

"Each child took all five forms of the modified-sentence

spelling test to permit the collection of information on the child's reaction to the spelling of each of the 300 words. The six forms of the recognition test were cycled randomly throughout the entire experimental population thereby requiring each child to take only one form of the recognition test." Each child took one form of the Biggy Spelling Recognition Test and five forms of the Biggy Modified-Sentence Test, the Niles Visual Test and the Niles Auditory Test.

The data were analyzed to determine:

1. The difficulty of items on the modified-sentence test.
2. The difficulty of items on the recognition test.
3. The reliability of the modified-sentence test.
4. The reliability of the recognition test.
5. The correlation between the recognition test and the modified-sentence test as an indication of the extent to which the two tests measured the same or different abilities.
6. The correlation between the modified-sentence test and visual test; between the modified-sentence test and the auditory test.
7. The correlation between the recognition test and the visual test; between the recognition test and the auditory test.
8. The correlation between the visual and auditory tests.
9. The difficulty value on each instrument for each of the 300 words used in the experiment.

Conclusions:

1. Approximately 16 percent of the population correctly

spelled just slightly better than half of the words presented.

2. The composite reliability of the five forms of the modified-sentence test when determined by the Spearman-Brown prophecy formula produces a value of .978 which indicates the modified-sentence test to be highly reliable.
3. An alarming degree of failure on common words is evident when the item analyses of Forms A through E of the modified-sentence test are reviewed.
4. A range of from three to ten words on each form were failed by 50% or more of the modal-age population.
5. Reliability coefficients computed by the Kuder-Richardson Formula (#20) (known to reduce the coefficient value by underestimation) obtained on the two independently drawn samples from each of the six forms of the Recognition Test indicated reliability of the highest type (.98).
6. A difference of 12 points in score on the recognition test in favor of the girls and of five words on the modified-sentence test again indicates the superiority of girls in spelling ability and ability to detect spelling errors.
7. The correlation between the two methods of measuring spelling ability (the recognition test and the modified-sentence test) is in the order of .6. Though the correlation is a comparatively high positive value it is not high enough to suggest that the two methods of measuring spelling ability are interchangeable.

8. Low correlations between the various tests indicate less relationship between the abilities measured by these tests than has most previous research.

Brooke, Barbara Ann. Comparison of Recognition and Recall of Word Elements as Methods of Measuring Visual and Auditory Perception in Spelling. Ed. M., 1947. 66p.

Problems:

1. To investigate the relative merits of recognition and recall tests in measuring visual and auditory perception, as these factors affect spelling ability.
2. To obtain any possible data concerning the frequency and position of error in the perception tests, and to see if any particular pattern of errors exists in either the incorrect recognition or the incorrect recall of the material used.

Materials and Procedure:

The writer used 180 children in grades four, five, and six.

The writer constructed tests consisting of configurations composed of common word elements - prefixes, suffixes and roots. Two forms of a recognition and two forms of a recall test were constructed, so the same set of configurations could be used to test both recall and recognition. This eliminated the possibility of practice affecting score if one set had been used and memory entered in. The population was divided in half with Form A of recognition and Form B of recall administered to one group with the second group taking Form B of recognition and Form A of recall. Both Visual and Auditory tests were built in this way. Repetition of word elements was guarded against.

The pupils' scores on the recognition and recall perception tests were correlated with the spelling results of the Inter-

showing on card.

Visual Recall- Teacher showed card and pupils wrote down what they remembered.

Auditory Recall- Same as visual recall, except teacher pronounced "word".

Flash cards were exposed three to five seconds. Each word was pronounced twice in immediate sequence. Care was taken to pronounce each sound in the "word" as clearly and as accurately as possible. Accent was placed in the part of the "word" where it seemed logical. Example: trib ment.

Carter, Bernadette R., The Construction Tests of Visual Perception, Auditory Discrimination and Kinesthetic Factors to be used in Diagnosing Inadequate Performance in Spelling.
Ed. M. 1941. 71p.

Problem: To determine whether tests constructed measuring visual perception, auditory discrimination and speed of handwriting can be used in the middle grades in diagnosing inadequate performance in spelling and to determine the relationship, if any, between the factors tested.

Procedure:

(1) The following tests were given to 180 pupils in grades five and six in the city of Attleboro, Mass.

- (a) Visual Perception of Words
- (b) Auditory Discrimination and Written Reproduction of Words Pronounced
- (c) Speed of Handwriting from Copy
- (d) Stanford Achievement Test of Spelling (Form D)
- (e) Metropolitan Achievement Test of Spelling (Form A)
- (f) Kuhlman Anderson Test of Intelligence

(2) A description of the constructed tests is as follows:

- (a) Visual Perception of Words.

In this test the child is asked to write on lined spelling paper the word which the examiner has just displayed for two seconds. The total number of words is 25.

Examples of some of the words used are as follows:

- (1) Variform)
- (2) Incumbent)
- (3) Exponent) etc.
- (4) Flotsam)
- (5) Gangliate)
- (6) Hydrostat)

(b) Auditory Discrimination For Words.

In this test the examiner pronounces a word and the subjects write it as it is pronounced. Words phonetically correct and those spelled correctly are scored correct. The total number of words is 25. Example of some of the words are as follows:

- (1) Synapse
- (2) Stimulus
- (3) Templet
- (4) Pilaster
- (5) Astrolabe
- (6) Bacillus

(c) Speed of Handwriting from Copy

In this test the pupils were required to copy a short poem of eight lines. Each child was given a typewritten copy to use at his desk. The children were asked to copy until the examiner directed them to stop. The score of the test equals the number of letters written in one minute, the highest possible score being 200. A copy of the poem is as follows:

"When the wind is in the east
Tis good for neither man nor beast
When the wind is in the north
The skillful fisherman goes not forth
When the wind is in the south
It blows the bait in the fishes mouth
When the wind is in the west
Then it is the very best."

(3) A correlation technique was used in comparing the tests to study the relationship between achievement and all other factors tested.

(4) Scatter diagrams were made to determine the importance of auditory, visual and kinesthetic factors in relation to spelling. An individual item analysis was made of 50% of the highest cases and 50% of the lowest cases to study the nature of the test items.

(5) The reliability of the visual and auditory test was deter-

mined by the split half Spearman Brown Method.

Conclusions:

(1) The correlations between the various tests with spelling achievement are as follows:

- (a) visual perception .6385
- (b) auditory discrimination .5725
- (c) mental age as measured by the Kuhlman-Anderson Intelligence test. 3220
- (d) Speed of handwriting from copy .0928

Mental age and speed of handwriting are not statistically significant.

(2) The auditory and visual factors are most important in discriminating between the very poor and the very good spellers.

(3) The scatter diagram shows:

(a) Of all pupils in the bottom 30% in achievement, 82% of them will lie in the low half of the auditory discrimination test. Of all those in the lower half in latter test, 74% of them are in the low half in the achievement test. It is evident therefore, that a poor speller is very likely (82 chances in 100) to be lower than average in auditory discrimination, and further those below average in auditory discrimination are also likely (74 chances in 100) to be below in spelling achievement. The relationship is not perfect, however. Of the pupils in the bottom 30% in spelling achievement 18% of them were above average in auditory discrimination. Such a discrepancy might partly be explained by the fact that such pupils may be low in visual discrimination.

(b) Similarly of those in the bottom 30% in achievement, 76% are in the bottom half in the visual perception test and 87% in the bottom six deciles. Of those in the bottom half in

spelling achievement test. Part of the 13% who are in the bottom three deciles in spelling achievement and who are in the four top deciles in visual discrimination, may be explained by the fact that they may be in the lower deciles in auditory discrimination.

(c) The data presented indicates that visual and auditory discrimination are important factors related to spelling achievement.

(d) The figures of the item analysis indicate that the tests of auditory discrimination and visual perception exhibit the proper range of difficulty, since none of the items in either of the tests were spelled entirely correctly and none were misspelled by all of the pupils.

Cavanagh, Francis James. A Comparison of Data from Selected Mental Tests. Ed.M. 1944. 57p.

Problem: A statistical comparison of I.Q.'s from some of the commonly used group mental tests when given to pupils whose grade placement levels are at the extremities of the test ranges.

Procedure: 1. The following tests were given to the following grades:

Grade IV

- a. Otis Quick Scoring Mental Abilities Tests, Alpha Test Form A, Gr. 1-4
- b. Otis Quick Scoring Mental Ability Test, Beta Test Form A, Gr. 4-9

Grade VIIIa

- a. The Henmon-Nelson Tests of Mental Ability Form A, Gr. 3-8
- b. The Henmon-Nelson Tests of Mental Maturity Form A, Gr. 7-12
- c. The Kuhlmann-Anderson Tests for Grades VII to VIII
- d. The Kuhlmann-Anderson Tests for Grades IX to Maturity

Grade IX

- a. Otis Quick Scoring Mental Ability Tests, Beta Test, Form A, Gr. 4-9
- b. Otis Quick Scoring Mental Ability Tests, Gamma Test, Form C, Gr. 9 through college.

2. The tests were administered in the following manner:

a. Grade IV: The Otis Beta test was given first. The Beta test was given a week before the Alpha Nonverbal. Five days passed between the giving of the verbal and non-verbal.

b. Grade VIII: The pupils of grade 8a were given the Henmon-Nelson 7-12 first and two days later the Kuhlmann-Anderson IX to

Maturity Test. A week later the Henmen-Nelson 3-8 was given and two days later the Kuhlmann-Anderson VII-VIII. The 8b section was given the Henmen-Nelson 3-8 test first and two days later the Kuhlmann-Anderson VII-VIII. A week later the Henmen-Nelson 7-12 was given and two days later the Kuhlmann-Anderson IX-Maturity.

c. Grade IX: The Beta test was given first and was followed by the Gamma test one week later. All tests in every grade were administered and scored by the author.

3. The mean, standard error, difference of means and critical ratio were computed for each test and each grade. The correlations were made, between tests in each grade to see their relationship.

4. Data was obtained on a total of 140 cases, 42 from Grade IV, 51 from Grade VIII and 47 from Grade IX.

Conclusions: 1. The Otis Alpha tests (the lower form) show a higher mean I.Q. (97.) than the Otis Beta Tests (the higher form) (92.4) when given to the same grade four population. The difference of the Means of these tests is noticeable (4.6). In the Grade IX tests the lower form test (Otis Beta has a higher mean (106.19) than the Mean of the higher form test (Otis Gamma) (105.46)). The differences in Means is .73. The results in the Grade VIII tests show the upper form tests to be higher in Mean value and Differences of Means. The Henmen-Nelson 7-12 test has a higher Mean (103.5) than the Henmen-Nelson B-8 (97.33) when used on the same population. The difference of means is 6.17. The Kuhlmann-Anderson IX to Maturity test also has a higher mean (101.9) than the Kuhlmann-Anderson VII-VIII test (96.45). The difference of the Mean is 5.436. While there is

some justification for educators to believe that the upper form tests will give higher I.Q.'s than the lower forms the results of this experiment only partially prove this theory.

2. The wide range shown by one test does not necessarily maintain itself in the result of other tests when comparisons of ranges are made. The number of intervals in the ranges of the Otis Beta and Otis Gamma Tests when given to the grade none population is five for each test. There is a noticeable difference in the number of intervals in the ranges of the Kuhlmann-Anderson VII-VIII and Kuhlmann-Anderson IX Maturity Tests as shown in the grade eight data. There are seven intervals in the range of the lower form and five in the range of the upper form. The Henmon-Nelson 3-8 Test has one additional interval of range than the Henmon-Nelson 7-12 Test as shown in grade eight results.

3. The great percentage of overlap indicates that while the individual tests do give a position of relative value they do not give an exacting discrimination value.

Cavin, Grace. An Analysis of Intermediate Grade Reading Tests.

Ed.M. 1940. 78p.

Problem: An analysis of 20 intermediate grade reading tests in respect to the following aspects: (1) abilities which tests measure, (2) nature of test items, (3) standardization of tests, (4) reliability of tests, (5) the extent to which the tests discriminate between grades.

Procedure: 1. The following tests were used in this study:

	<u>Grades</u>
(a) Detroit Reading Test, Test III	4-6
(b) Durrell-Sullivan Reading Achievement	3-6
(c) Gates Reading Survey	3-10
(d) Gates Silent Reading Tests	3-8
(e) Haggerty Reading Examination	6-12
(f) Ingram Clark-Diagnostic Reading Int.	4-8
(g) Iowa Silent Reading Test	4-9
(h) Los Angeles Elementary Reading Test	3-4
(i) Metropolitan Intermediate Reading Test	4-6
(j) Monroe Standardized Silent Reading Test 1	3-5
(k) Nelson Reading Test	3-9
(l) New Stanford Reading Test	2-9
(m) Pressey Diagnostic Reading Tests	3-9
(n) Progressive Reading Tests Elem.	3-6
(o) Public School Achievement Tests, Reading	3-8
(p) Sangven-Woody Reading Test	4-8
(q) Shark Tests of Reading Comprehension	3-6
(r) Thorndike-McCall Reading Scale	2-12
(s) Unit Scales of Attainment Div. 1	3-4
(t) Unit Scales of Attainment Div. 2	5-6

2. The 20 tests were analyzed according to the five aspects stated in the problem. The source of information was the manuals accompanying these tests.

3. The method followed in analyzing the abilities which the tests measure, was to compile a list of the functions of each test as stated in the manuals. A table was set up list-

ing the abilities stated and the number of tests measuring these abilities.

4. In analyzing the test items, vocabulary tests were considered first and tests of paragraph comprehension second. They were discussed from three standpoints - length, content, and comprehension checks. Tables were set up showing percentage of vocabulary, kinds of words, and distribution of words in each test. Tables were set up showing number of paragraphs, timing of paragraphs, and content of reading material for each test. A list of the number of questions in each test and the different kinds of comprehension checks used was also drawn up.

5. The standardization of the tests considered the nature of the population and the number of cases used. Information was obtained from the manuals and was wholly inadequate for the purpose of analysis. Tables were set up listing type and size of population used for each test.

6. Information concerning the reliability of the tests was obtained from the manuals and was also incomplete. Tables were set up showing the number of reliability coefficients reported for each test, the population from which these coefficients were derived, and any other reliability data reported by tests.

7. The step-up intervals for each grade was found in the following manner: The number of raw points necessary to achieve each grade level was secured from the grade norms of a test. The raw score intervals between grades 3 and 4, 4 and 5, etc. were computed and averaged. The resulting figure is the average step-up interval between grades on a test.

Conclusions: 1. The stated abilities of what the 20 tests measured differ widely, both in respect to number and kind. The announced function of a test is not necessarily an adequate description of what the test measures. For example, in the Progressive test organization means the ability to make use of references. In the Sangren-Woody test organization means arranging shuffled events in their correct sequence. The Los Angeles Elementary test measures 14 abilities which their tests measure.

2. It is impossible to determine whether the norms reported for tests are representative. This is because insufficient information is published in the manuals. This is true in regard to the reliability coefficients also.

3. Tests differ greatly in respect to their power of discriminating grades. For example, in the Detroit Test III the average interval is 2.9 and the Monroe test is 1.9. Measuring instruments possessing so little grade discrimination would seem not to be too reliable.

4. If the 20 tests under consideration show such great variability in respect to the factors for which they were analyzed, it cannot be concluded that they are measures of reading ability per se.

Cogen, Anna M. A Comparison of The Vocabularies Of The Grade One and Two Books of the Alice and Jerry Series, the Curriculum Foundation Series, the Today Work Plan Book Series and the Gates Primary Reading Standardized Test. Ed. M. 1950. 55 p.

Problem:

To compare the vocabularies of the above tests and texts to determine how well each test sampled the words of the basal reader and what per cent of the words in the test were in the basal readers.

Materials:

1. The texts included four pre-primers, a primer, a first reader, a second reader of the Alice and Jerry series, three pre-primers, a primer, a first reader, and two second readers of the Curriculum Foundation series, and two pre-primers, a primer, first reader, and second reader of the Today's Work and Play series.
2. The test was the Gates Primary Reading Standardized test.

Procedure:

1. All the words of the test were recorded and alphabetized.
2. Alphabetical lists were made for the basal readers, each book of the three series being checked separately.

Major Findings and Conclusions:

1. The texts varied widely in the number and percent of words used.

Cozanove, Frances S., et al.* The Construction and Analysis
of Tests to Measure the Knowledge of Letter Names in Grades
One, Two and Three. Ed.N. 1953. 38p.

Problem: To develop and analyze measures to test the knowledge of letter names in Grades 1, 11, and 111.

Procedure: 1. Five group tests and one individual test were given. The group tests included: (a) Written Identification of Letters (the auditory factor was of prime importance in the construction of this test). Care was taken in the dictation in order to eliminate auditory confusion of such letters as b and v, m and n, f and v. (b) Identification of Letters in Groups from Dictation. (This test consisted of 28 items with five letters in each item. The letter combinations were carefully analyzed as to contiguity and likeness and difference in form.) (c) Identification of Letters in Groups from Flash Cards. (This was a test to determine the child's knowledge of the letter name through the association of the upper and lower case letters, e.g., when H was flashed, the correct response was h.) (d) Identification of Lower Case Letters in Unfamiliar Words When Upper Case Letters are Flashed. (This test consisted of 28 words taken from Laidlaw Basic Readers and Gates Basic Readers. When a card was flashed, the correct response was the corresponding lower case letter found in the unfamiliar word). (e) Identification of Lower Case Letters in Unfamiliar Words Through Dictation. (f) An Individual Oral Inventory of All Upper and Lower Case Letters was given. (The letters were arranged in such a way

that no memory pattern of the child's could be used.)

2. The testing was done on 609 children in Grades 1, 11, and 111 in 21 classrooms in 17 different communities. The group was heterogeneous including children from low, middle and upper socio-economic communities. The intelligence quotient was measured by the Otis Quick Scoring Test.

3. Frequency distribution tables were set up showing the variation of scores in each test in each grade. The mean and standard deviation were found for each test.

4. A reliability study showed the correlations between an individual oral inventory test and the knowledge measured by five group tests.

Conclusions: 1. The children in Grades 11 and 111 had a good knowledge of letter names as measured by all the instruments.

(a) The mean score for Test I in Grade 11 was 50.98, letters with a S.D. of 2.16. In Grade 111, 51.25 letters with a S.D. of .88.

(b) The mean score for Test II in Grade 11 was 24.60 with a S.D. of 2.78. In Grade 111, 24.90 with a S.D. of 1.55. (c) The mean

score for Test III in Grade 11 was 27.40 letters with a S.D. of 1.68. In Grade 111, 27.40 letters with a S.D. of 2.08. In

Grade 111, 27.40 letters with a S.D. of 2.12. (e) The mean score for Test V in Grade 11 was 27.38 letters with a S.D. of

1.31. In Grade 111, 27.64 letters with a S.D. of 1.76. (f) The mean score for Test VI in Grade 11 was 27.40 letters with a S.D.

of 1.24. In Grade 111, 27.50 letters with a S.D. of 1.15.

2. All of the tests showed a wide range in Grade 1 and fairly high reliability. For example, in the Individual Test the scores in Grade 1 ranged from 0-52 (53 was the upper limit).

In Test 11 they ranged from 0-26, in Test 111 0-28, Test IV, 0-28, Test V, 0-28 (29 was the upper limit), Test VI from 0-28 (29 was the upper limit).

3. It would appear that children in Grade 11 and 111 do have a good knowledge of letter names and still may have difficulty in reading. In Grade 11, 176 cases out of 215 were at the top in Test 1. In Grade 111, 158 out of 184 were at the top interval. In Test 11, 186 cases out of 215 were at the top interval in Grade 11. In Grade 111, 155 out of 184 were at the top interval. In Test 111, 194 out of 215 were at the top interval in Grade 11. In Grade 111, 157 were at the top interval. In Test IV, 172 out of 215 cases were at the top interval in Grade 11. In Grade 111, 162 out of 184 cases were at the top interval. In Test V, 191 out of 215 cases were at the top interval in Grade 11. In Grade 111, 170 out of 184 cases were at the top interval. In Test VI, 190 out of 215 cases were at the top interval in Grade 11. In Grade 111, 172 out of 184 cases were at the top interval.

* Hill, Grete A., Stevens, Freda C.

Daniels, Katherine H. An Evaluation of Certain Informal Reading Tests. Ed. 1940. 70p.

Problem: 1. Informal reading tests to discover the child's reading level were given by nine teachers, a supervisor, and an assistant to 246 children in grades one, two and three.

2. The three basal readers that were used in the informal testing were "The Child Development", "Alice and Jerry" and "The Children's Own Readers" series. Four were taken from each series and the children had not seen any of them. An attempt was made to choose materials of equal difficulty, but since it was impossible to equate them perfectly a schedule was set up so that no one series was used consistently by one group.

3. Standardized tests, both oral and silent, were given by the supervisor. At least two oral and two silent tests were given to each child in order to have a composite score yielding more accurate data. The following standardized tests were used in the experiment:

- Grade I. Metropolitan Achievement 1-3 Form A, Gates Primary 1-3 Form 2, Gray Oral Reading Paragraphs, Durrell Oral Reading Paragraphs.
- Grade II. Metropolitan Achievement 1-3, Form A, Gates Primary 1-3, Form 2, Durrell-Sullivan Reading Capacity 3-6, Gray Oral Reading Paragraphs, Durrell Oral Reading Paragraphs.
- Grade III. Metropolitan Achievement 1-3, Form A, Gates Silent Reading Test 3-8, Form 2, Durrell-Sullivan Reading Capacity 3-6, Gray Oral Reading Paragraphs, Durrell Oral Reading Paragraphs.

4. In the informal testing the first and last page of

a basal reader was used; if the child made more than seven errors on a page he was given a series lower to read. Uniform questions were asked after each testing and a check list of errors was provided (taken from Durrell Analysis of Reading Difficulty, pg.2).

5. In the standardized tests the silent reading tests were given as a group, the oral tests individually.

6. Inter-correlations were made between the ratings given by the three teachers (Room Teachers, Supervisor, Assistant) and between the composite test score on several standardized tests and the teachers' ratings.

7. Frequency distribution tables were set up showing the scores and ratings given by the Room Teachers, Supervisor and Assistant in grades one, two and three.

Conclusion: The correlations were as follows:

- (1) Correlations between room teachers' ratings and a composite test score:

Grade I	.82	\pm	.02
Grade II	.90	\pm	.01
Grade III	.81	\pm	.02

- (2) Correlations between classroom teachers' ratings and a composite test score:

Grade I	.82	\pm	.02
Grade II	.90	\pm	.01
Grade III	.78	\pm	.03

- (3) Correlations between supervisors' ratings and the composite test score:

Grade I	.90	\pm	.01
Grade II	.90	\pm	.01
Grade III	.88	\pm	.015

4. The assistant with less experience showed lower correlations between her ratings and the composite score:

Grade I	.78	+	.01
Grade II	.65	+	.05
Grade III	.74	-	.03

5. The comparisons between the room teachers' ratings and the composite test score showed a tendency for the room teacher to rate the children higher than the composite test score.

Higher rating	48.9%
Equal rating	35 %
Lower rating	16.1%

6. The median scores differ for the informal tests and for the standardized tests in all three grades.

Grade I.	Informal test medians	1.9-1.9-1.6		
	Standardized test medians	2.0-2.0	1.7	1.8
Grade II.	Informal test medians	2.9 2.9 3.1		
	Standardized test medians	2.7 3.2 3.4	3.2	
Grade III.	Informal test medians	4.1 3.6 3.9		
	Standardized test medians	3.3 3.7 3.5	3.736	

7. There is a lack of uniformity in the results checked by all the three examiners in the check list of errors.

Errors checked by one examiner	969
" " " two examiners	289
" " " three "	65

8. The classroom supervisor checked more errors than the supervisor or assistant.

Room teachers	563
Supervisor	470
Assistant	290

De Loura, Bertha Amy. A Study of the Predictive Value of Intelligence and Aptitude Tests to Reading Achievement in Grade One. Ed.M. 1949. 71p.

Problem: An attempt to evaluate the Pintner-Cunningham Primary Mental Test, Form A as a test of intelligence and the Marion Monroe Aptitude Test as a test in reading readiness, in determining the predictive measure of intelligence and readiness to reading achievement.

Procedure: 1. The two tests, Pintner-Cunningham and Marion Monroe, were administered to 267 children in the first grade in a community in the Boston area.

2. The following May the Metropolitan Achievement Test was given to the same children who had taken the other tests.

3. A correlation technique was used, the scores on the reading tests were correlated with scores on achievement tests.

4. Correlations were computed also, between the sexes and reading achievement and between the different intelligence groups and reading achievement. (I.Q.'s were obtained from the Pintner-Cunningham tests Range 53-163).

Conclusions: 1. I.Q. seems to have an important relation to readiness. The coefficient of correlation was .76. This has a forecasting efficiency of 35%.

2. A knowledge of child's mental age is an aid in estimating his probable success or failure in reading, as the

correlation between mental age and reading achievement has a predictive value of 10% better than chance.

3. The relationship between I.Q. and achievement did not seem to be statistically significant. The coefficient of correlation was .28-5% forecasting efficiency.

4. There was little value for prediction shown in relation of readiness to achievement. The coefficient of correlation was .28 with a forecasting efficiency of 4%.

5. The tests did not favor one sex more than the other. Critical ratio of scores between tests was 1.8- of no statistical significance.

De Wolfe, Barbara Ellen, A Comparison Of The Vocabularies Of The Easy Growth In Reading Series And Ten Standard Reading Tests. Ed. M. 1948. 100p.

Problem: To compare the vocabularies of ten standardized reading tests with the vocabulary of the Easy Growth In Reading Series from the pre-primer to the second level to determine

- 1) What percentage of the words on each test are words in the basal reading system.
- 2) What percentage of the total vocabulary in the basal reading series is tested by each of the tests.
- 3) Which one or ones of the ten tests would give the most reliable and valid score for a pupil whose basal reading series is the Easy Growth In Reading.

Material: The tests used in the study are: 1) American School Achievement Tests (Primary 1), 2) Detroit Reading Test (Grades 1 and 11), 3) Detroit Word Recognition Test (Grade 1), 4) De Vault Primary Reading Test (Grades 1 and 11), 5) Gates Reading Tests (Grades 1 and 11), 6) Lee Clark Reading Test (Grades 1 and 11), 7) Los Angeles Primary Word Recognition (Grades 1 and 2), 8) Metropolitan Achievement Tests (Primary 1 and 11, 9) Reilly Primary Reading Test (Grade 1) 10) Unit Scales of Attainment (Grades 1 and 2).

Procedure: The vocabulary of each form of the ten tests was taken from the test and listed alphabetically on paper folded in columns with a letter of the alphabet at the top of each column. The same procedure was followed for the basal series. Both lists were compared and checked with a check list. The data was analyzed and tables were set up showing the percentage

of words on each test which were in the basal reading series, the percentage of the total vocabulary in the series which is tested by each test, and the most reliable and valid test in the series.

Conclusions:

- 1) The percentage of words in the 10 tests which are basal reader words ranged from 76.2 to 31.1 percent.
 - a) Form 1 of the De Vault Reading Test has the highest percentage, 76.2 percent.
 - b) The Unit Scales of Attainment, has the lowest, 31.1 percent.
- 2) The percentage of the total basal vocabulary tested by each test is very low, ranging from 38.3 percent to 8.7 percent.
 - a) The highest percentage is in the Unit Scales of Attainment which is 38.3 percent.
 - b) The Los Angeles Primary Word Recognition has the lowest percentage, 8.7 percent.
- 3) Not one of the tests includes enough of the vocabulary of the basal reading series.

Dodge, Harriet Eva. A Comparison of the Stanford-Binet Mental Age and the Hearing Comprehension Ability of Dull Children.
Ed. M. 1940. 33p.

Problem: This study attempts to answer the following questions:

1. Is the difference between Stanford-Binet mental age and hearing comprehension as determined by the Durrell-Sullivan Reading Capacity Test statistically significant?

2. Is there a difference in accomplishment quotients when mental age and hearing comprehension are used as criteria of ability?

3. Are there statistically significant sex differences in mental age, hearing comprehension, reading age and achievement quotient?

4. Do older children improve in hearing comprehension and remain retarded in mental age?

Procedure: 1. Complete data was obtained on 558 cases from the metropolitan cities, the east, (Lynn, Mass.; New Haven, Conn.; and Rochester, N.Y.).

2. The tests used in the study were the Stanford-Binet Form "L" and Durrell-Sullivan Reading Capacity and Achievement.

3. Since the results of the Binet tests secured from New Haven and Rochester were old, current mental ages were derived by using the straight line prediction. All derived mental ages were checked and rechecked. The scoring of tests and tabulation of results were checked.

4. Frequency distribution tables were set up for: chronological age, mental age, reading age, hearing comprehension, and hearing comprehension mental age differences. This was done as combined data and data split by sexes. Mean standard error and standard error of the mean were figured for each of these.

5. Where desired for purposes of comparison, the standard error of the difference and the critical ratio were computed.

Conclusions: 1. The difference between mental age and hearing comprehension as determined by these tests is statistically significant as critical ratios of 14.73 and 21.95 indicate.

2. In the Lynn group the difference between accomplishment and capacity is statistically significant if reading capacity is measured by hearing comprehension: the critical ratio is 12.69. In the Rochester group the difference is almost statistically significant; the critical ratio is 2.86. If accomplishment is rated against capacity determined by Binet mental age the differences are not statistically significant; critical ratios are 1.16 and 1.59.

3. The achievement quotient when determined by Reading-Age-Hearing Comprehension is .82 for the Lynn group and .95 for the Rochester group. When determined by Reading Age-Mental Age the achievement quotient is 1.01 for the Lynn children and 1.03 for the Rochester children. Thus it would seem that dull children work up to and beyond capacity if it is measured by hearing comprehension.

4. Sex differences in Chronological Age and

Mental Age: The girls were found to be older chronologically and younger mentally than the boys. (Mean mental age of boys 8-10-50 and of girls 8-6-14). Differences in the boys and girls' chronological ages and in the boys and girls' mental ages were not statistically significant.

5. Sex differences in Hearing Comprehension: The three groups of data show the mean hearing comprehension age of the boys to be above that of the girls. In the Lynn group this difference is statistically significant; critical ratio is 4. In New Haven and Rochester the data is not statistically significant, but show a trend in the same direction; critical ratios 2.27 and 1.55. Sex differences in Achievement show that girls achieve more early to capacity than do the boys, although the differences are not statistically significant.

6. The relationship of hearing comprehension to mental age seems to remain constant at different ages. Children grow in mental age as rapidly as in hearing comprehension.

Doherty, Mary Louise. Comparison of Ten Standardized Tests with the New Work-Play Books. Ed. M. 1948. 81p.

Problem: To compare the vocabularies of ten standardized reading tests with the vocabularies of the Ginn basal readers, preprimer through second reader levels.

Materials:

1. The following tests: American School Achievement Tests, Detroit Reading Test, Detroit Word Recognition Tests, DeVault Primary Test, Gates Primary Reading Tests, Lee-Clark Reading Tests, Los Angeles Primary Word Recognition Test, Metropolitan Achievement Tests, Reilley Reading Test, and Unit Scales of Attainment.
2. The books used were: Off We Go, Jim and Judy, Down Our Street, and We Grow Up.
3. All the words from each test were listed in alphabetical order.
4. The words from the basal readers were listed and alphabetized, by level. Derivatives were counted as new words except for plurals formed by adding s. Proper names and interjections were counted as new words.
5. Graph paper was used for making the check lists of words, with the word lists of the basal readers placed diagonally along the left side of the paper and the names of the tests across the top of the paper.
6. One person read the list of words from the test and another checked each word appearing in the list for the

basal readers.

Major Findings and Conclusions:

1. The percentages found of the test vocabulary were generally low. The DeVault Primary Reading Test, Form I ranked the highest with 80%. The second highest was the Detroit Reading Test, Form A, with 75%. The Unit Scales of Attainment ranked lowest with 30%.
2. Percents of basal reader vocabulary were extremely low. The two forms of the Unit Scales of Attainment for Grade I, First Half, ranked highest of the basal reader vocabulary. Form A was 42 per cent; Form B, 40 per cent. Three tests were as low as nine per cent. They were: Detroit Word Recognition, Form D; Los Angeles Primary Reading Tests, Forms I and II.
3. None of the tests seemed to include a vocabulary which was representative of the vocabulary of the Gates basal readers.

Euphrosyne, Georges et al*. Evaluation of Exercises Designed to Increase Children's Use of Descriptive Words. Ed. M.

1951 160 p.

Problem: To make an evaluation of exercises designed to increase children's use of specific words.

Procedure: 1. The experiment was conducted with 108 third grade children in a community near Boston.

2. The children were divided into two groups, the experimental and the control. Two tests were given to both groups to determine the number of descriptive words in the child's speaking vocabulary.

3. The first test simply determined the number and quality of words that an individual could give in two minutes; the second test consisted of words given in response to an observation of a picture. After the tests were given to both groups a teaching period of six weeks was given to the experimental group. The lessons consisted of 24 exercises designed for increasing children's use of descriptive words. The control group had their regular language program with no particular emphasis on descriptive words.

4. For a one week period prior to the April vacation re-tests were given to both groups. All the testing was given individually.

5. The tests were scored and then compared to see if there were any statistical differences between groups.

Conclusions: 1. Descriptive Words: The combined results of picture and two minute stimuli showed a difference of 31.51% in favor of the experimental group. The critical ratio was 3.53 in favor of the experimental group.

2. Descriptive Phrases: The data shows a difference of 21.70% in favor of the experimental group. The critical ratio was 2.79.

3. Non-Descriptive Words: An analysis of the data shows a difference of 49.83% between control and experimental groups. The critical ratio of 5.18 was in favor of the control group.

*MacPartlin, Anna I.

Farr, Carol V. Construction and Evaluation of Tests of Rate of Learning in Reading for Children in the First Grade. Ed.M. 1942. 55p.

Problem: To construct and evaluate tests of rate of learning in reading for children in the first grade.

Procedure: 1. Learning rate was defined as the number of words a child has learned in one controlled reading period as indicated by (1) Identification Check, (2) Picture Check, immediately after the controlled teaching period.

2. A group of words was selected from the International Kindergarten List that was part of the child's hearing vocabulary but not reading vocabulary. They were carefully checked against Primary Reading Lists and basal texts and were given individually to 25 children taken from the lower third of the classes, to make sure they were part of the speaking vocabulary but not reading vocabulary of the children. As a result of the test .66 words were selected.

3. Two types of tests were constructed and administered: (1) Identification Check: A series of words similar in initial consonants and endings including word taught. (2) Picture Check: Words taught were placed in a column and on either side were placed illustrations of these words, supplemented by extra pictures to lessen guessing. The test was to draw a line from word to correct picture.

4. Included in the data for each child was a readiness test given in September and an achievement test given in

June. These tests were correlated to see the relationships between readiness and achievement and their predictive value. Correlations were also made between the achievement tests and learning rate tests, and the readiness tests and learning rate tests. The following tests were used: Meyropolitan Readiness Test, Gates Reading Test and Metropolitan Achievement Test. Complete data was obtained on 50 children.

Conclusions: 1. The comparison of the results of the rate of learning tests when different number of words are used yielded the following: (a) The mean score on the five word test is 23.14 compared with 30.54 on the seven word test and 42.98 on the ten word test. 77% of the words taught in the five word test were learned compared with 72% in the seven word test and 71% in the ten word test. There is a greater per cent of learning when five words are taught. The mean scores are higher on the seven and ten word tests, but the per cent of learning is less.

2. No conclusions can be drawn about the relative difference of the Picture Check or the Identification Check. When all the Identification and Picture means are totalled there is only one point difference.

3. Correlations between the Rate of Learning tests and the two Reading Readiness tests are as follows: (a) The coefficient of correlation $.45 \pm .07$ of the Rate of Learning and the Metropolitan Readiness is low. It is concluded that a high rate of learning does not necessarily follow a high readiness score. (b) The correlation $.48 \pm .06$ of the Rate of Learning and the Gates Readiness Tests reveals only a slight relationship between the learning capacity and the readiness to read.

Foss, Gertrude Mae, Language Comprehension Skills of Mentally Retarded Children. Ed. M. 1938. 37p.

Problem: To discover the relationship between various language skills of mentally retarded children.

1. To determine the effect of a foreign language background on hearing comprehension scores and on reading achievement scores.

2. To discover the frequency of special reading difficulties by determining the number of children who are reading below their mental age and the number whose reading achievement is below their hearing comprehension; and to discover the correlation between MA's and reading achievement and hearing comprehension and reading achievement.

3. To discover the extent to which the understanding of spoken language influence the child's reading achievement.

Procedure: Stanford-Binet; The Durrell-Sullivan Reading Capacity; The Durrell-Sullivan Achievement Test.

"The fundamental assumption of these tests (the last 2) is that serious reading disabilities can be discovered by revealing discrepancies between the child's understanding of spoken language and his understanding of the printed word."

The Durrell-Sullivan Capacity test is divided into two parts: Word Meaning-consists of 70 words which are tested by 14 groups of pictures to measure the child's hearing vocabulary; Paragraph Meaning- The child demonstrates his understanding of the para-

graph read to him by marking the number of the picture which illustrates the answer to the question asked him.

Conclusions: The paired technique was used in this study.

1. The mean score for the English Speaking group in hearing vocabulary was 45.9 and the mean score for the Foreign Speaking group was 42.8. This showed a difference of 3.2 in favor of the English Speaking group.

2. Effect of foreign language in the home on Hearing Comprehension showed a difference of 3.5 in favor of the English Speaking group. The mean score for the Foreign Speaking group was 35.9 and the mean for the English speaking group was 39.5.

3. There was no significant difference on the achievement test.

4. Reading Achievement in relation to hearing Comprehension showed a difference of 12.6 in the mean scores in favor of the English Speaking group. The mean for the English Speaking group was 39.8 and the mean score was 27.2 for the Foreign Speaking group.

Fox, Ruth E. A Comparison of Ten Standardized Reading Tests with the Curriculum Foundation Series. Ed. M. 1948. 52p.

Purpose: To find the percentage of words included in the test and in the basal series, as well as the percentage of the vocabulary of the basal series tested.

Materials: From the Scott Foresman Curriculum Foundation Series the three pre-primers, the primer, the first reader, and the first and second level second readers were used. The following ten reading tests were also used in the study:

1. American School Achievement Tests. Forms IA, and IB for grade one; IIA, IIB, and IIC for grade two.
2. Detroit Word Recognition Test. Forms A, B, C, D for grade one.
3. Detroit Reading Test. Forms A and B both for grades one and two.
4. De Vault Primary Reading Test. Form I for grades one and two.
5. Gates Primary Reading Tests. Forms I, II, and III all for grade one and first half of grade two.
6. Lee - Clark Reading Test. Primer Forms A and B for grade one, and First Reader Forms A and B for grades one and two.
7. Los Angeles Primary Word Recognition Test. Forms I and II both for grades one and two.
8. Metropolitan Achievement Tests. Primary I Battery Forms R and S for grade one, and Primary II Battery Forms R and S for grade two.

9. Reilley Primary Reading Test. Forms A and B for grade one.
10. Unit Scales of Attainment. Forms A and B for grade one first half, forms A and B for grade one last half, forms A and B for grade two first half, and forms A and B for grade two last half.

Procedure: The vocabulary lists were made by alphabetizing the new words presented in each book. The words from the tests were then checked against the five basic reader vocabulary lists.

Findings: The percent of words in both the tests and the reading series varied from 33 to 73 for the first grade tests, 50 to 81 for the second grade tests, and 59 to 91 for the first and second grade tests. The percent of the Curriculum Foundation Series vocabulary tested ranged from 12 to 37 in the tests for the first grade, from 9 to 34 in the tests for the second grade, and from 8 to 28 in the tests for the first and second grade. The tests which sampled the highest percent of vocabulary of the Curriculum Foundation Series were the Reilley Primary Reading Test Form A for the first grade, the American School Achievement Test form B for the second grade, and the Gates Primary Reading Tests Forms 1 and 2 for the first and second grades.

Gilman, Alice. Mental Test Variations Among Dull Children.

Ed.M. 1941. 104p.

Problem: The problem is twofold: (1) to discover and evaluate the mental test variations among dull children. (2) to discover the frequency of reading disability among the same group, by using two criteria of ability; the Durrell-Sullivan Hearing Capacity Test and the Stanford-Binet individual examination, each in comparison with the Durrell-Sullivan Reading Achievement Test.

Procedure: 1. The following tests were used in the study:

- (a) Stanford Revision of the Binet-Simon Tests
- (b) Durrell-Sullivan Reading Capacity Test
- (c) Durrell-Sullivan Reading Achievement Test
- (d) California Test of Mental Maturity Primary Battery
- (e) Pintner-Durost General Ability Tests, Elementary Battery, Form A.

2. Complete data was obtained on 189 special class children whose chronological ages ranged from seven years-ten months to eighteen years-one month. The average age of the group was thirteen years-three months. The I.Q.'s of the group range from 47-95 with an average I.Q. of 71.2.

3. The tests were given in six to eight testing periods. Derived mental ages were checked as well as the scoring and tabulation of test results.

4. Frequency distribution tables were set up showing: chronological ages; I.Q.'s; Binet mental ages; Hearing Capacity scores, which were treated as mental ages; reading ages derived from the Durrell-Sullivan Reading Achievement Test;

total Pintner-Durost Test mental ages; non-reading scores; reading scores; total California Test of Mental Maturity mental ages; non-language scores; language scores.

5. Correlations between the tests and subtests were found by the Pierson Product Moment method of correlation. Means, standard error of the means, standard deviation, standard error of the differences, critical ratios were obtained for each of the above.

Conclusions: 1. The results of the study showed a very marked variation in the degree in which the tests agreed, each with the other, showing a range of mean mental ages from 97.4 on the Pintner-Durost General Ability Test to 119.1 on the Durrell-Sullivan Hearing Capacity Test. Other mean mental ages were 107.3 on Binet, 102.5 on Reading Achievement, 101.6 on the California Test of Mental Maturity. Coefficients of correlation ranged from $.79 \pm .027$ between Durrell-Sullivan and the California Test to $.55 \pm .51$ between Pintner-Durost and Binet.

2. There was not a significant difference between the verbal and non-verbal sub sections of either the California or Pintner-Durost Tests for this population. Pintner-Durost non-reading mental age 99.2; reading mental age 96.1; California language test mean mental age 102.3; non-language test 100.9.

3. The critical ratios were much higher when group test scores were compared with the Hearing Capacity mental age mean, than when they were compared with Binet mental age mean. (Pintner-Durost 11.24; California 8.66; Binet 6.02).

4. When the group test scores were compared to

reading achievement critical ratios ranged from 8.16 on the Hearing Capacity Paragraphs to .001 between Reading Achievement and the Language Form of the California Test of Mental Maturity.

5. There was further disagreement when each test was taken in turn as a standard of criterion for reading achievement. On Binet 31% would be considered reading disability cases, when retardation is assumed to be one full year below estimated ability. On Hearing Capacity, 64% would be considered reading disability.

6. Taking both Binet mental ages and Hearing Capacity as criteria of reading ability, statistical evidence showed that they fell short of their estimated capacity on both measures. The achievement quotient of Reading Age-Hearing Comprehension was .86, and for Reading Age-Mental Age, .96.

Goggin, Dorothea D., et al.* Construction and Statistical Analysis of Group Auditory Tests for Grades One Two and Three. Ed.M. 1953. 39p.

Problem: To construct and validate as many different measures as possible to test the ability to hear the consonant sounds at the beginning, middle and end of words.

Procedure: 1. Two group auditory tests and one individual inventory were constructed for grades I,II, and III. A third group auditory test was taken from a thesis by Conway and others and used also.

2. Test 1 has two parts, testing initial and final consonants. A sample item taken from the test is as follows: "Today we are going to play a game with some letters." (Draw a picture of a ball on the board. Put these letters under it: c b f l.) The teacher says the word in the picture, children put X under the beginning letter. The same procedure is followed for the last letter.

3. Test 2 is an auditory test taken from thesis of Conway and others. A sample item from the test is as follows: "Today we are going to play a game with some letters. We are going to see how well you hear sounds. Listen to the word I say, "Mimic"; Put a ring around the letters you hear in mimic".
c r i e m p Unfamiliar words were chosen in order to test auditory discrimination.

4. Test 4 was designed to measure auditory perception through the use of context clues. A sample item is as follows:

"Today we are going to play a game with some letters. We are going to see how well you hear sounds. Listen to this story: "Humpty Dumpty sat on a wall, Humpty Dumpty had a great _____. The child is supposed to think of the word missing in the story and then think of the letter it begins with and put a cross under it. c m b f s

5. Test 4 is an individual inventory of the twenty-one consonants. The letters are arranged unalphabetically and the child is supposed to recognize them.

6. The tests administered by classroom teachers were given to 234 first grade children, 217 second grade children and 187 third grade children in metropolitan Boston. The I.Q.'s of these children were slightly above normal (Grade I - 106, Grade II - 111, Grade III - 105).

7. After administering and scoring the tests they were statistically analyzed. The frequency distribution, mean and standard deviation were found for each test.

8. Correlations were made with each of three tests with Test 4 the individual inventory and the standard error was found on these correlations.

9. An item analysis was done on ten cases randomly selected around the mean to compare children's ability and to see which letters were found most difficult.

Conclusions: 1. All tests were good measures for grade one. The correlations of the group test to the individual were the following:

Test 1	.64
" 2	.43
" 3	.89

All tests were fair for grade two. The following correlations were:

Test 1	.52
" 2	.55
" 3	.55

All tests were poor for grade three:

Test 1	.43
" 2	.45
" 3	.34

2. There was little relationship between test two measuring ability to recognize sound in speech, and test four measuring single letters. The correlations with test four was

Grade 1	.43
" 2	.55
" 3	.45

This test had the lowest correlations.

3. Test three, giving sounds from context clues, showed the highest correlation with test four:

Grade 1	.89
" 2	.55
" 3	.34

4. The item analysis showed: (a) a general increase in the knowledge of the sound of the letter with the increase in grade level, e.g., the letter m in grade one, two out of ten got it right; in grade two, four; and in grade three, seven. This follows with the other letters also. (b) The final sound of the letter was consistently the most difficult. (c) Although the sound was not known individually it sometimes was known in the word. (d) Certain letters continue to present difficulty throughout the three grades. They were found to be l, n and r.

* Hughes, Margaret Theresa; Wall, Rita Elizabeth.

Greenleaf, Edith Eleanor. An Evaluation of Visual Perception Tests for Predicting Success in First Grade Reading. Ed.M. 1936. 34p.

Problem: To determine the validity of the methods used in the Newton Reading Readiness Tests measuring visual discrimination and to study the relation between the perceptual abilities measured in these tests and various types of reading achievement.

Procedure: 1. In May and June 1935 all the kindergarten children of the Newton Public Schools were given the five Newton Reading Readiness Tests and either the Pintner-Cunningham Primary Mental Test or the Kuhlmann-Anderson Test for Grade One.

2. In March, 1936 those same children were given four Primary Reading Tests based on the Pennell and Cusack Readers.

3. The following data was obtained for 471 cases used in this study: (a) Chronological Age from test taken in 1935, (b) Mental Age from 1935 test, (c) I.Q. obtained from either the Pintner-Cunningham or Kuhlmann-Anderson test, (d) five raw scores on the Newton Readiness Tests, (e) four raw scores on the Newton Primary Reading Tests.

4. In addition to this, between March and April a flash and pronunciation test was given using 30 items used in the basic readers. Complete data was obtained on 104 cases.

5. The total scores for 471 cases on Primary Reading Tests based upon work covered in the Children's Own Readers by Pennell and Cusack were correlated with the scores on each of the

five Newton Reading Readiness Tests (tests of visual discrimination and perception). The correlation between the total scores on both sets of tests were also computed. Correlations were also made on scores of 104 children on the Pronunciation Test and the Reading Readiness Test.

Conclusions: 1. The Newton Primary Reading Test is unsatisfactory for the purpose of statistical evaluation because the tests were too easy. With a maximum score of 124, 200 of the 471 children got scores between 155-124 causing the distribution of scores to be skewed. This prevented reliable computation of correlations. Since a standard reading test was not available it was decided to use the Pronunciation test as a measure of achievement.

2. The results of this investigation give no conclusive evidence as to the particular value of using as measures of Reading Readiness Tests, designs or words, comparison or memory, or selection of a wrong element after visual discrimination between similar work forms. There can be no conclusion as to which type of test is better, whether word and letter forms are better than designs. The correlation between tests of design comparison is 0.39 compared to word comparison 0.37, but in memory words have a higher correlation. Words 0.45, Designs 0.28. The same conflicting evidence is seen for comparison and memory. The test of visual memory has a higher correlation than comparison of words, but memory of designs has a lower coefficient than comparison. Test 3 in the Reading Readiness Test measures the ability to select a wrong element after visual discrimination between similar word forms. It had a correlation of 0.27 with the Pronunciation Test (of no statistical significance.)

3. There is some evidence that the ability to compare designs (/0.39) is more predictive of later success in reading than the ability to remember designs (0.28) or to compare words (0.37).

4. The visual perception test, testing the ability to remember word forms, has a correlation of 0.45 with the pronunciation test and appears to be the type of test that has the greatest relation to reading achievement.

5. The ability to persist and work independently is probably an important factor in reading success. Reading Readiness Test (1), Design and Comparison (2), Word Comparison, require this ability. Design Comparison had a correlation of 0.39 with Pronunciation Test and Word Comparison correlation of 0.37.

6. The evidence to the conclusion that the tachistoscope test is probably a reliable measure of achievement in reading. The correlation of composite score on Reading Readiness Test and Flashed Word Test was 0.46.

Haggerty, Earl J. An Evaluation of Certain Tests Used as Measures of Reading Capacity. Ed.M. 1940. 74p.

Problem: A study of the comparisons of various tests of mental capacity (Stanford Revision of the Binet-Simon Tests, Kuhlman-Anderson Intelligence Tests, Henmon-Nelson Tests of Mental Maturity, Durrell-Sullivan Reading Capacity Test) with reading achievement (Gates Silent Reading Tests).

Procedure: 1. The data for this study was obtained from 233 fourth grade children whose chronological ages ranged from eight years two months to eleven years nine months. The mental capacity of the children was determined from individual Stanford-Binet tests given in kindergarten. The I.Q.'s ranged from 75 to 145 with a mean I.Q. of 113.

2. The Durrell-Sullivan test was given in 1940 at mid-semester for the purpose of comparison with other tests, intelligence quotients were computed. The Kuhlman-Anderson test was given when the children were in the third grade. The Henmon-Nelson test was given in 1940 at mid-semester. The Gates Test was given in 1939 by the school psychologist during the first semester.

3. Since the results of the Stanford-Binet, Kuhlman-Anderson and Gates tests were obtained from tests at different intervals, current mental ages and reading ages were derived in order to obtain a more valid comparison. The straight line prediction was used in deriving these scores.

4. Frequency distributions were set up for chronol-

ogical ages, mental age, reading age and I.Q. Means, standard error of the means, standard error of the differences and critical ratios were computed.

5. Correlations between tests were found by the Pearson product method of correlation.

Conclusions: 1. Comparison of measures of reading capacity and reading achievement gave the following results:

(a) The mean for the Henmon-Nelson was 130.2 \pm .967 as compared to 134.3 \pm 1.313 of Gates tests. This is not statistically significant.

(b) The mean for the Binet, 129.1 \pm .836 as compared to 134.3 for the Gates gave a critical ratio of 3.25. This was statistically significant.

(c) The mean for the Kuhlman-Anderson, 127.1 \pm .823 as compared to 132.1 \pm 1.410 for the Gates, gave a critical ratio of 3.13 showing a statistically important difference.

(d) The mean for the Durrell-Sullivan was 124.2 \pm .862 as compared to 134.3 \pm 1.313 for the Gates, and gave a critical ratio of 6.31 showing a statistically significant difference.

2. Correlations between measures of reading capacity and achievement were as follows:

(a) Henmon-Nelson and Gates .667 \pm .037

(b) Durrell-Sullivan and Gates .517 \pm .049

(c) Kuhlmann-Anderson and Gates .448 \pm .060

(d) Binet and Gates .214 \pm .036

3. In general, significant differences were found between reading capacity and reading achievement as measured in this study as shown in the critical ratios numerated above except

in the Henmon-Nelson ratio which is not significant.

4. Except in the case of the Stanford-Binet, where the degree of relationship was low ($.214 \pm .036$) there seemed to be a marked degree of relationship between reading capacity and reading achievement as measured in this study.

Haskell, Barbara A. et al.* The Relationship of the Knowledge of Letter Names and Reading Achievement in Grade One. Ed.M. 1952. 31p.

Problem: To discover the relationship between the knowledge of letter names and reading achievement in Grade 1.

Procedure: 1. A battery of three tests was constructed to measure the abilities of first grade children in the following areas:

- (a) Visual Perception of Letters: This test consisted of 42 items, each of which contains four letters selected on the basis of similarities in configuration.
- (b) Recognition of Letter Names: This test consisted of 31 items, each of which contained four letters. These were selected on the basis of similarities in configuration, and frequency of letter difficulty.
- (c) Knowledge of Capital and Lower Case Letters: This test consisted of 44 items, each of which contained four letters selected on the basis of similarities in configuration and frequency of letter difficulties. In addition to these tests the following standard tests were given:
 - (1) The Otis Quick Scoring Intelligence Test
 - (2) The Detroit Word Recognition Test

All testing was given in February by classroom teachers and scored by authors.

2. The number of first grade children tested was 639 and these represented farming, industrial and residential

Massachusetts communities. The mean chronological age of the group was 6-6 mos. and the mean I.Q. was 105.80.

3. A correlation technique was used in making comparisons between reading and visual perception, intelligence quotient, sex differences.

Conclusions: The results of the data indicate:

1. There seems to be little relationship between the knowledge of letter names and reading achievement.

(a) The correlations range from .31 to .54.

(b) The highest correlation was between association of upper and lower case letters and reading ($r = .54$).

2. There seems to be a definite relationship between the intelligence quotient and reading achievement in favor of the children in the 110 and above group. The critical ratio is 4.90.

3. The girls were superior in reading achievement and in all of the factors studied.

(a) The mean score of the girls in reading was 12.70; for the boys, 10.51.

(b) The mean score of the girls on letter association was 35.62 letters; for the boys, 33.91.

(c) The mean score of the girls in the knowledge of letter names was 27.55; for the boys, 25.84.

(d) The mean score of the girls in the visual test was 40.90; for the boys, 40.42.

Hughes, Alice A. An Analysis of the Vocabulary of Two Standardized Reading Tests in Relation to the Vocabulary of Three Reading Systems. Ed. M. 1950. 61p.

Problem: To compare the vocabularies of each part of two standardized reading tests with the vocabularies of three basal reading systems, preprimer through second reader levels.

Materials:

1. The tests used were American School Achievement, Forms I-A and II-A and Metropolitan Achievement, Forms I-R and II-R.
2. The basal reading systems used were the Today's Work-Play Books, the Reading Foundation Series, and the Curriculum Series.

Procedure:

1. The words in each test in each battery were listed and alphabetized. All variants were listed as separate words.
2. Alphabetized lists of the vocabulary of the reading books were made.
3. Check sheets were prepared in order to compare the test and reader vocabularies. A separate check sheet was made for each test in each battery. Test words were listed in alphabetical order at the left. Three columns were made beside the words, with the name of a reading system at the top. As each word on the reading

system list appeared in the exact form on the test list, it was checked off.

Major Findings and Conclusions:

1. The per cent of words included in the two standardized reading tests and in the three basal reading systems was not very high. The per cent ranges from 26 to 45 in Grade One and from 27 to 54 in Grade Two.
2. The per cents of Grade One words in common with the tests were quite evenly divided among the three basal reading systems. The Grade Two vocabulary of the Curriculum Foundation Series had the highest per cent of words in common with the two tests among the Grade Two vocabularies of the three basal reading systems.
3. The per cent of the vocabulary of the three basal reading systems which are included in the two tests was very low. The per cent ranged from six to twelve in Grade One and from six to thirty-two in Grade Two.
4. The percents of Grade One words sampled by the two tests are quite evenly divided among the three basal reading systems.

Hurlburt, Lydis Delpha, The Relative Value of Recall and Recognition Techniques for Measuring Precise Knowledge of Word Meaning, Nouns, Verbs, Adjectives. Ed. D. 1949, 159 p.

Problem: To construct crucial and incisive knowledge of word meanings for junior and senior high school in order to obtain information about:

(1) The relative value of recall and recognition techniques for meanings.

(2) The significance of different parts of speech for testing word knowledge.

Procedure:

(1) After making an extensive study of much literature about vocabulary building, reading, and the allied phases of language, a list of many techniques and methods for testing vocabulary ability was made.

(2) From this list the following techniques were selected for building a battery of tests that would adequately measure a high school student's precise knowledge of word meanings.

Recognition

Multiple Item True-False	Synonyms Noun, Verb, Adjective
Multiple Item True-False	Antonyms Noun, Verb, Adjective
Multiple Item True-False	Applied Meanings Nouns, Adjectives

Recall

Completion	Synonym	Noun
Completion	Synonym	Verb
Completion	Synonym	Adjective

(3) Since the number of variables in the multiple item true-false tests made it impossible to equate such a form of recognition test with the recall tests, it was decided to hold the three multiple item true-false tests and prepare recognition matching tests with the identical items of the recall-completion tests that would be equivalent in every detail. Thus, the test lay out became:

Nouns (Synonyms)

Recall-completion form

Recognition-matching form

Verbs (Synonyms)

Recall-completion form

Recognition-matching form

Adjectives (Synonyms)

Recall-completion form

Recognition -matching form

(4) For the words in all the multiple item true-false tests Webster's Dictionary of Synonyms was the source. A master list of 2,000 words was made and then rated for usage according to the Thorndike-Lorge count. A hundred words were chosen by taking a random sampling of every twentieth word on the master list. A similar procedure was followed for the antonyms. For the multiple item true-false test of applied meanings a master list of 500 words was gathered from Mawson's Roget's Thesaurus of the English Language in Dictionary Form. The test list of key words was selected in the same manner as for the antonym and synonym tests. For the recall-completion tests words were selected from Webster's Dictionary of Synonyms and Mawson's arrangement of Roget's Thesaurus. Separate lists were made of all the acceptable words for each part of speech; noun, verb, adjective.

(5) The final form consisted of separate tests of 115 items each for nouns, verbs and adjectives. Each test was prepared in two forms with identical items: recall-completion and recognition-matching. The recall form was designed to measure an individual's facility in summoning words of synonym relationship for use in speaking and writing. The recognition form was designed to measure an individual's ability to identify and attach precise meaning to words.

(6) The experiment was conducted in the ninth and eleventh grades in the high school for a suburb of Boston during the month of May, 1948. The testing program included:

- (1) California Short-Form Test of Mental Maturity (Advanced)
- (2) Vocabulary-Noun (Recall and Recognition Forms)
- (3) Vocabulary-Verb (Recall and Recognition Form)
- (4) Vocabulary-Adjective (Recall and Recognition Forms)

The statistical analysis was made on 402 students in the school. Coefficients of correlation and critical ratios were computed to determine the degree of relationship between factors or the statistical significance of difference in achievement for test forms and separate tests.

Conclusions: 1. Comparison of recall and recognition test forms.

A. Achievement in raw scores

1. Total vocabulary score (combined nouns, verbs, adjectives.)

(a) The average high school student of Grades IX. and XI. was able to recall and write only 45% of the words he was able to recognize.

(b) Bright students were able to recall and write approximately 52% of the words they were able to recognize.

(c) Dull students were able to recall and write about 40% of

the words they were able to recognize.

2. Separate tests

The average high school student of Grade IX, and XI, was able to recall and write approximately:

- (a) 60% of the noun words he was able to recognize.
- (b) 42% of the verb words he was able to recognize.
- (c) 32% of the adjective words he was able to recognize.

B. Correlations.

1. The coefficients of correlations between achievement on recall and recognition test forms was positive (Recall and Recognition Gr. IX, 23.037, Recall and Recognition Gr. XI, 24.25) but only moderately high. Therefore it is considered that:

- (a) The two vocabulary test forms have only a limited number of factors in common.
- (b) Achievement on either test form cannot be used to predict achievement on the others with any degree of certainty.
- (c) The correlation between achievement on recall and recognition test forms is no higher for the bright students (c.r. $.61 \pm .07$ Gr. IX.) ($.52 \pm .10$ Gr. XI.) than it is for the total population ($.62 \pm .04$ $.60 \pm .05$)

2. The correlation between recall or recognition test form and mental age was only moderately high (recall .48 for Gr. IX, and .58 for Gr. XI.) and the recognition test form correlated as well with mental age as the recall form did (recognition .47 for Gr. IX, and .52 for Gr. XI.)

11. Comparison of achievement on separate tests.

A. Achievement in raw scores.

(1) The highest raw scores occurred for nouns and the lowest for adjectives. (nouns-recognition Gr. 1X. 61.525, Gr. XI. 71.190, recall Gr. 1X. 34.208, Gr. XI. 46.004) (adjectives-recognition Gr. 1X. 34.945, Gr. XI. 47.285, recall- Gr. 1X. 10.842, Gr. XI. 17.142)

(2) The mean raw score for nouns was three times greater than the mean raw score for adjectives on both recall and recognition test forms.

(3) The mean raw score for verbs was two times greater than the mean raw score for adjectives on both recall and recognition (verbs recognition Gr. 1X. 55.7, Gr. XI. 68.710; (verbs recall Gr. 1X. 22.456, Gr. XI. 30.316) (adjectives recognition Gr. 1X. 34.945, Gr. XI. 47.285; recall Gr. 1X. 10.842, Gr. XI. 17.142).

(4) The mean raw score for nouns was 1.5 times greater than for verbs on the recall test form (nouns Gr. 1X. 34.208, Gr. XI. 46.004; verbs Gr. 1X. 22.465, Gr. XI. 30.316.) but on the recognition test form the difference in raw scores for nouns and verbs was negligible. (nouns Gr. 1X. 61.525, Gr. XI. 71.190; verbs Gr. 1X. 55.7, Gr. XI. 68.71)

B. Correlations

(1) Separate tests and mental age

The correlation between the separate tests and mental ages was approximately the same as for test forms or a single test form and mental age. The highest correlation of .43 occurred for nouns and the lowest of .43 occurred for adjectives.

(2) Among separate tests

The correlation between the separate tests ran somewhat higher than the correlations between the test forms or between either test form and mental age. The highest correlation of .79 occurred

for nouns and adjective recognition form. The lowest correlation of .53 occurred for verbs and adjectives recall form. Correlations were consistently higher for the recognition form of the separate tests than the recall form. Recall-noun Gr. 1X. .757- Verb .606 Adjective .583. Gr. XI. Noun .652, Verb .538, Adjective .5598. Recognition Gr. 1X. Noun .794, Verb .715, Adjective .794 Gr. XI. Noun .657, Verb .716, Adjective .657.

III. Sex difference in vocabulary scores.

There were slight differences in mean raw scores for boys and girls. Those differences were consistently in favor of the boys. Gr. 1X. Boys, 221.54, Girls 212.82, Gr. XI. Boys 238.28, Girls 269.72

IV. Grade differences in vocabulary scores.

Mean raw scores for both forms of each test run appreciably higher for Gr. XI. than for Gr. 1X.

V. Test reliability.

Recall and recognition forms with identical phrase definitions for test items as used in this study are reliable measures of precise knowledge of word meaning for nouns, verbs, adjectives.

Jack, Barbara. A Comparison of the Vocabularies of Ten Standard Achievement Test in Reading with the Vocabulary of the Ginn Basic Readers. Ed. M. 1948. 63p.

Purpose: To determine what percentage of words included in the tests are words in the basal system, and what percentage of the basal vocabulary is tested.

Materials: From the Ginn Basic Readers the three pre-primers, the primer, the first and second level second readers were used. The following ten reading tests were also used in the study:

1. American School Achievement Test. Forms IA and IB for the first grade; IIA, IIB, and IIC for the second grade.
2. Detroit Word Recognition Test. Forms A, B, C, and D for grade one.
3. Detroit Reading Test. Forms A and B both for grades one and two.
4. De Vault Primary Reading Test. Form I for grades one and two.
5. Gates Primary Reading Tests. Forms I, II, and III all for grade one and first half of grade two.
6. Lee-Clark Reading Tests. Primer Forms A and B for grade one and First Reader Forms A and B for grades one and two.
7. Los Angeles Primary Word Recognition Test. Forms I and II both for grades one and two.

8. Metropolitan Achievement Tests. Primary I Battery Forms R and S for grade one, and Primary II Battery Forms R and S for grade two.
9. Reilley Primary Reading Test. Forms A and B for grade one.
10. Unit Scales of Attainment. Forms A and B for grade one first half, forms A and B for grade one last half, forms A and B for grade two first half, and forms A and B for grade two last half.

Procedure: The words listed as "new words" at the back of each book were arranged alphabetically. The words from each pre-primer and second reader were then alphabetized with the other words on the same level, with the result that there was one alphabetical list for each level -- pre-primer, primer, grade one, and grade two. The words from each form of the tests were then arranged alphabetically. The lists of reader vocabulary and test vocabulary were compared.

Findings: The percent of words in both the tests and the reading series varied from 22 to 67 for the first grade tests, 41 to 76 for the second grade tests, and 47 to 80 for the tests for the first and second grade. The percent of the Ginn Basic Readers vocabulary tested ranged from 10 to 40 in the tests for the first grade, from 20 to 32 in the tests for the second grade, and from 9 to 32 in the tests for the first and second grades. The tests which sampled the highest percent of vocabulary of the Ginn Basic Readers were the Unit Scales of Attainment grade one, half one or two, form A or B,

and the Reilley Primary Reading Test form A for the first grade; Unit Scales of Attainment grade two, half one, form A or B, and the American School Achievement form B for the second grade; and the Gates Primary Reading Tests forms I, II, and III from the tests for both the first and second grades.

Junkins, Kathryn M. The Construction and Evaluation of Exercises for Developing Visual Discrimination in Beginning Reading.

Ed. M. 1940 148 p.

Problem: To build exercises for visual discrimination and to evaluate their effect upon beginning reading as measured by the following factors:

1. Effect on visual discrimination
2. Effect on learning rate
3. Effect on word recognition test scores
4. Effect on auditory discrimination

Procedure: 1. Materials were set up in outline form covering a period of six weeks, or thirty teaching lessons.

2. Each lesson was limited to ten minutes.

3. The following types of exercises were included in the plans:

- a. Matching single letters
- b. Matching pairs of letters
- c. Matching parts of words
- d. Matching words in isolation
- e. Finding words in context
- f. Matching ideas expressed by words
- g. Making of letters
- h. Making of words

4. The words in the study were chosen from the Gates Primary Word List: "Gates, A.I.: "A Reading Vocabulary for the Primary Grades (Revised and Enlarged) Bureau of Publications, New

York City, 1935" -- and from words of high frequency in the International Kindergarten Union -- "A Study of the Vocabulary of Children Before Entering First Grade ", Distributed by the International Kindergarten Union, Washington, D. C.

5. In order to evaluate these exercises some basis for measurement had to be used so the author undertook the construction of tests which would measure:

- a. Visual discrimination
- b. Auditory discrimination
- c. Learning rate

These tests were given at the beginning and closing of the experiment.

Test 1. Visual discrimination: The test consisted of twenty-four groups of letters and words.

- (a) The first eight groups were letters arranged in difficulty from letters easily distinguished to those more difficult.
- (b) The last sixteen groups were words placed in groups of five progressing in difficulty from easily distinguished ones to more difficult ones.

Test 2. Auditory discrimination: A group test was built; the Monroe Reading Aptitude Auditory test (1) suggested the form.

Test 3. Learning rate: To test the learning rate eight words which had not been taught as reading words were selected from the Kindergarten Vocabulary List.

Other tests administered were

- (a) Detroit Advanced First Grade Intelligence, Published 1928 by World Book Company, New York to determine the mental ages of the groups.
- (b) Detroit Word Recognition Test Form A, Published 1925 by World Book Company, New York to determine

what effect this training might have upon reading. These groups, for the most part, were made up of children of foreign backgrounds from the same district,

Major Findings and Conclusions: 1. In the visual discrimination test the experimental group was superior to the control group although the control group had a better mean mental age. This seems to prove that the ability to see likenesses and differences improves by specific teaching.

2. The visual discrimination exercises improved the rate of learning new words. This was evidenced by the ratio of the control group 2.76 as compared with 4.41 for the experimental group.

3. In the Detroit Word Recognition test the experimental group was superior to the control group.

4. The gains made in the group auditory test by the control group were larger than those made by the experimental group.

Kennedy, Helen Ursula. An Evaluation Of Ten Standard Achievement Tests In Reading In Comparison With The D. C. Heath Basic Readers. Ed. M. 1948. 44p.

Problem: To compare the vocabulary used in ten Standard Achievement Tests in reading with the vocabulary of the Reading for Interest Series (Pub. by D. C. Heath Co.,) from the pre-primer level through the second reader to determine:

1. What percentage of words in each test are included in this basal system.
2. What percentage of the basal vocabulary in this system is tested.

Materials: The ten tests analyzed were the following: 1.) American School Achievement Test, (Grades 1 and 11), 2.) Detroit Reading Test, 3.) Detroit Word Recognition Test (Grade 1.), 4.) Gates Primary Reading Test, 5.) Lee-Clark Reading Test (Grade 1 and 11), 6.) Los Angeles Word Recognition Test (Grades 1 and 11), 7.) Metropolitan Achievement Test (Grade 1 and 11), 8.) Reilley Primary Reading Test (Grade 1), 9) Unit Scales of Attainment (Grades 1 and 11), 10) De Vault Reading Test.

Procedure: The new words in each book were listed in alphabetical order. As each word was transcribed on the list it was checked off the book list to avoid duplication. The same procedure was followed for the test words. Five people working in co-operation took two tests each and listed the words in alphabetical order. The two lists were compared and tables were set

up showing the percentage of words included in the tests that were in the basal readers and the percentage of basal vocabulary tested in the basal system.

Conclusions:

1. The tests vary greatly in the number of words included in the tests, which are words in the basal system.

a. The highest percentage of test vocabulary which is basal reader vocabulary is 69 percent in the Detroit Word Recognition Test, Form A and 67 percent in Detroit Reading Test, Form B and the De Vault Reading Test, Form 1.

b. The lowest percent is 39 in the Unit Scales of Attainment.

2. The number of words in the basal system which are tested is generally low.

a) The highest was 35 percent in the American School Achievement Test, Form 11.B.

b) The lowest was 8 percent in both the Los Angeles Primary Word Recognition, Form 1 and the Detroit Word Recognition, Form D.

Kent, Helen H. An Evaluation of a Test for Predicting Reading Success in Grade 1. Ed.M. 1948. 66p.

Problem: To determine how far the Thurstons' Tests of Primary Mental Abilities measures reading readiness and its relationship between reading achievement.

Procedure: 1. Three communities were chosen from the Greater Boston area to represent high, middle, and low socio-economic levels.

2. Frequency distribution tables containing the Primary Mental Abilities quotient scores for the high, middle, low socio-economic groups for boys and girls were set up. Using the method of critical ratio, the level of significance for sex differences was determined.

3. In January, 1948 the Test of Primary Mental Abilities was given to 229 children of the first grade.

4. Late in March, 1948 the Metropolitan Achievement Test for reading Primary 1, Battery Form R (revised) for Grade 1 was given to the same group of children.

5. The coefficient of correlation was computed between the Primary Mental Abilities total score and reading achievement and intercorrelations were computed for the subtests and reading achievement to determine which had the greater efficiency for prediction of reading success.

Conclusions: 1. A strong relationship between the Primary Mental Abilities scores and reading achievement is not completely borne

out by the coefficient of correlation. The coefficients of correlation between the Primary Mental Abilities and Reading Achievement .49. This forecasting efficiency of 13% better than chance has moderate value for prognosis. This does give an indication that a knowledge of a pupil's intelligence quotient is an aid in predicting a child's success, but that the predictive value of the Primary Mental Abilities test was limited as a prognostic instrument of reading achievement.

2. Of the subtests in the Primary Mental Tests, Quantitative and Space were the most significant in determining a child's success in first grade reading. The coefficient of correlation between Quantitative and reading achievement was $\underline{.43}$ with a standard error of $\underline{.028}$. The coefficient of correlation between Space and reading achievement was $\underline{.45}$ with a standard error of $\underline{.30}$. These correlations indicated a predictive value of about 10% better than chance. The motor test had the lowest correlation $\underline{.27}$ with a standard error of $\underline{.18}$ giving only a 4% better than chance prediction.

3. There were no statistically significant differences between the results obtained for the girls and boys on the tests. The mean score for the 132 boys was 105 and 105.2 for the 126 girls. The mean difference between the two was .02 in favor of the girls.

4. There were statistically significant differences between the results obtained for the quotient of the high socio-economic group and the middle socio-economic group. The mean intelligence quotient of the high group was 110.35 as compared with 100.05 for the middle group. The critical ratio of 5.0

shows a statistically significant difference. The arbitrary assumption that the low socio-economic groups was below the middle group is questionable. The mean intelligence quotient of the middle group was 100.05 as compared to 104.63 for the low group. Critical ratio of 2.12 shows difference to favor lower group.

Kiernan , Anna Constance. The Construction and Evaluation of an Auditory Test of Word Analysis. Ed.M. 1950. 47p.

Problem: To construct a group auditory test of word analysis and to check its validity against standardized tests of reading achievement and intelligence.

Procedure: 1. An auditory test of 67 items was constructed. It includes the following:

- (a) Part I. Initial Single Sounds. Examiner: "What letter do these words begin with?" bacon-beauty-biscuit. j-g-h-k-b-a-l. The child encircles the correct letter. Letters tested : b-g-r-h-c-s-p-d-j-n-k-t-l-w-y.
- (b) Part II. Beginning Blends and Prefixes. Examiner: "Put a line around the letters which tell the beginning sounds of the words I say." thunder, thumb, thief. wh-sh-th-st-br. Blends and prefixes tested : th-wh-tr-st-sh-dr-pl-sm-tw-con-in-re-per.
- (c) Part III. Isolated Single Sounds. Examiner: "Listen to the sound. I shall say it three times." b-b-b; j-g-h-k-b-a. Child puts a circle around correct letter. Sounds tested : b-g-h-c-l-v-m-f-t-p.
- (d) Part IV. Medial Sounds. Examiner: "This time we are going to listen for beginning letters and for other letters." Mark the letters you hear: villain; s-v-l-a-c-f-n. Medial sounds tested: l-b-g.
- (e) Part V. Endings. Examiner: "Put a line around the ending

letters. Sometimes you will put a line around one letter and sometimes around more than one letter. Listen to the way the word ends : flood-hid-load; z-t-s-d-r-c. Final single sounds tested: d-k-g-p. Endings tested: ed-ing-er-ous. Final blends: ch-al-sh.

(f) Part VI. Phonograms. Examiner: "You will put a line around a whole word that rhymes with my words." band, hand, sand; cab, rust, land, salad, fort. Phonograms tested : and, it, ap, esp, arn, ent, ab, ut, orn.

2. The test was given to 96 second grade children in two city schools. It was given in the first week in March in each school. The Otis Mental Ability Test and the Stanford Primary Reading Tests were given the same week.

3. These tests were correlated with the auditory test to see their relationship. The test was also analyzed for the frequency and distribution of the scores of each part of it.

Conclusions: 1. The test indicated some relation to reading achievement. On the whole, good readers received high scores and poor readers low scores. The mean score of the high group was 61.34. The critical ratio of 3.80 is statistically significant. The mean score of the low group was 47.18. The critical ratio of 6.53 is statistically significant.

2. When the auditory test scores are related to the intelligence scores there is no definite pattern. The mean score of the high Otis group was 60.14 with a critical ratio of .883. The mean score of the middle group was 52.77 with a critical ratio of .995. The mean score of the low group was

50.61 with a critical ratio of 1.11. None of them are statistically significant.

3. The best part of the test is Part VI, Phonograms. This test has a fair range and a good distribution of scores. The highest score was 9 and the median was 5.71. There were as many children receiving a score of 3 as there were with 6. As many received 1 as 7.

4. In every other part the test seems to be too easy. Part I, 49 out of 96 children received perfect scores with the median 14.52 out of a possible score of 15. Part II, the median is 11.15 out of a possible score of 13. The interquartile range of 4.34 shows that it is better than Test 1, but still too easy. Part III, the median 9.63 and the interquartile range 1.27 was the smallest of any test. Part IV, 81 children out of 96 received scores of 7 or above out of a possible score of 9. The median score was 8.07. The interquartile range was 8.81. This was the best range of any test. Part V, the median was 9.17. The highest possible score was 11.

Macgregor, Helen Denison. An Analysis of the Vocabulary of Ten Selected Reading Tests and the Vocabulary of the Reading Foundation Series, "the Alice and Jerry Books." Ed. M. 1948. 120p.

Problem: To find out how well the vocabulary of ten selected standardized reading tests compare with the vocabulary in the Reading Foundation Series, The Alice and Jerry Books.

Materials:

1. The books analyzed were: Skip Along, The Sky, The Door, On a Hill, The New Day In and Day Out, The New Round About, and The New Friendly Village.
2. The standardized tests were: American School Achievement Test, Detroit Reading Test, Detroit Word Recognition Test, DeVault Primary Test, Gates Primary Test, Lee-Clark Primer Reading Test, Lee-Clark First Reader Los Angeles Primary Word Recognition Test, Metropolitan Achievement Test, Reilley Primary Test, and Unit Scales of Attainment.

Procedure:

1. The vocabularies listed at the back of each of the pre-primers were combined and arranged in alphabetical order and counted.
2. The vocabulary of the primer, first reader, and second reader was each arranged in alphabetical order and counted.

3. The same procedure was used with the tests, words from each form of each test being arranged in alphabetical order.
4. Tables were set up. As each word on the test list appeared on the basal list a check was made opposite the word under the test and form in which it appeared.

Major Findings and Conclusions:

1. The tests varied greatly in the number of words tested and in the sampling of the vocabulary.
2. The DeVault Primary Test, Grade One and Two, Form I, had the highest number of words tested while the Unit Scales of Attainment, First Half Grade One, Form A, had the lowest number of words tested.
3. The Unit Scales of Attainment, First Half Grade One, Forms A and B, had the greatest sampling of vocabulary while the Los Angeles Primary Word Recognition Test, Grade Two, Form B, had the lowest sampling of vocabulary.
4. No test had an adequate sampling of vocabulary for the Reading Foundation Series.

McAuliffe, Anita Serena, An Item Analysis Of Hurlburt Vocabulary Tests. Ed. M. 1948. 168p.

Problem:

1. The Hurlburt vocabulary test is devised for the purpose of measuring ninth and eleventh graders facility in using synonyms both in speaking and writing. The test consists of 115 items for each part of speech (nouns, verbs and adjectives.)

2. The analysis was made on 200 tests, 100 in each of grades nine and eleven. Fifty of the top scores and fifty of the lowest were studied.

3. The items were analyzed for:

- a) Discrimination between grades nine and eleven.
- b) Distribution of item difficulty
- c) Distribution of item validity
- d) Diagnostic possibilities of the item

Conclusions:

1. Nouns The item difficulty of the nouns had a range of .5 to .89. Items 7,20,60,67,83,85,92,93,113,114 have no grade discrimination for the ninth graders had higher scores than the eleventh. Items 1,5,11,17,21,33,4,48,53,59,72,73 and 80 showed a wide difference between the ninth and the eleventh grade.

2. The range of item validity was from .70 to 1.

3. Items 1,2,3,4,6,7,21,30,31,34,36,45,47,49,65,86,102 obtained 70 or more attempts by students.

4. Verbs Such items as 29,32,35,50,52,54,75,80,86,87, 93,96,98,101,108,114 were found to be the most difficult. The range of validity response was from 57 to 1.

5. Adjectives Adjective meaning had a very low correct response but had a very high attempts scores. The item analysis shows difficulty in the upper 50 cases as well as the lower 50. Many items had poor validity. Highest score made was 28 and the lowest was 0. This does not show a wide discrimination between grades nine and eleven.

McNiff, Margaret Ellen et al.* The Construction of a Vocabulary Test for the Intermediate Grades. Ed.M. 1953. 95p.

Problem: To construct two equivalent forms of an intermediate-grade vocabulary test to measure word recognition and word meaning, stressing a wide sampling of vocabulary and maintaining an economy of time and space.

- Materials:
1. Roget's Thesaurus
 2. Webster's Dictionary
 3. Thorndike's A Teacher's Word Book
 4. Rinsland's Basic Vocabulary
 5. Boston University Clinic List.

Procedure: 1. Each of the authors developed a separate test, using the classification technique, building it with 300 words, presented in 20 15-item exercises or in all 1200 items. The words included adjectives, adverbs, verbs, and nouns with no particular pattern as related to the number per category, and excluding any words ambiguous in meaning, archaic, etc.

2. The children were to respond by numbering each item to match the figure which preceded one of the categories.

3. Meanings were checked in Roget's Thesaurus.

4. Frequencies were checked with Thorndike's list, Rinsland's list, and the B. U. clinic list.

5. The tests were administered, untimed, to a total of 431 children in grades four, five, and six and a few from grade seven in three suburban communities of varying socio-economic backgrounds.

6. I.Q. information on the children was obtained.

7. Final test forms were developed, using the items showing approximately a 12% inter-grade increase in success. Each consisted of 10 exercises ranging from easiest to most difficult.

Major Findings and Conclusions: The completed test met the requirements of the problem. It consisted of a wide sampling of vocabulary at a minimum of time and space.

* Kelly, Dorothy Marie, MacDonald, Carolyn Elizabeth, McCarran, Mary Theresa.

Nason, Doris E. A Comparison of the Vocabularies of the Grade One and Two Books of the Learning to Read Series and Ten Standardized Reading Tests. Ed. M. 1948. 88p.

Problem: To compare the vocabularies of the texts and test above to find out what per cent of the words in the test were in the basal reader and how well the test sampled the words of the reader.

Materials:

1. The vocabularies from the Learning to Read series, first and second pre-primer, primer, first reader, and three second grade readers.
2. The following tests: American School Achievement, Unit Scales of Attainment, Detroit Word Recognition, Detroit Reading, Lee-Clark Primer and First Reader, DeVault Primary, Metropolitan, Los Angeles Primary Word Recognition, Reilley, and Gates.

Procedure:

1. Alphabetical vocabulary lists were made from the readers and tests.
2. Words appearing on reader list and on a particular test list were checked in columns beside the vocabulary lists.
3. Only if the form of the word on the test was it checked.

Major Findings and Conclusions:

1. Great variation was found in the number, per cent of

words, and extent of sampling of vocabulary when the comparisons were made.

2. The tests had a smaller per cent of words in common with the Grade I vocabulary than with the Grade II vocabulary.
3. The tests sampled the basal vocabulary for Grade I a little better than for Grade II.
4. The tests containing the largest per cent of vocabulary were the DeVault, Detroit Word Recognition and Detroit Reading.
5. None of the tests studied was suitable as an accurate measure of reading achievement in Grades I and II.

Nichols, Augusta Matilda. The Construction and Use of a Group Test for the Analysis of Spelling Difficulties. Ed. D., 1947. 220p.

Problem: The study has as its objectives:

1. The construction of a diagnostic group test in spelling.
2. The use of the test in detecting spelling difficulties and in locating their cause.

Materials and procedure: The experimentation was carried on in an industrial New England city. The Nelson Instructional Reading Tests which were administered yearly showed on the whole average or above average achievement. Form A of the Nichols Test was administered to 1615 pupils in grades three, four, five, and six. Form B was administered one month later to 1524 children.

The two final forms contained the following sub-tests.

- Sub-test I- Spelling Achievement
- Sub-test II- Proof Reading
- Sub-test III- Word Meaning
- Sub-test IV- Handwriting
- Sub-test V- Visual Discrimination
- Sub-test VI- Auditory Discrimination

Test items in both Form A and Form B were selected from the same source and were as nearly as could be determined of the same degree of difficulty.

A brief description of each test follows:

Sub-test I- Spelling Achievement-- Each form of this test contains 98 words, 14 on each grade level, grades two to eight

inclusive. Each grade spells 14 words on a lower level, 14 on its own level, and 28 on two higher levels. The spelling words in this test were selected from the list of Gates, Durrell, The Buckingham Extension of the Ayres' Spelling Scale, and from the first 6,000 words of the Buckingham Dolch Free Association Study. Each word on grade levels two through six was found on at least two of the first three lists, while the words for the seventh and eighth grade levels were found on the Gates and Free Association lists. The location of each word was determined by the grade placement given it on the Gates' list. Other phases of the spelling problem have been included in this test. A means of diagnosing spelling difficulties is provided by including words containing silent letters, words which may be spelled phonetically, homonyms, and words representing certain spelling rules:

1. Drop the final e before a suffix beginning with a vowel.
2. When final y is preceded by a consonant, change y to i before adding any suffix that does not begin with i.
3. Q's are always followed by u
4. The sound of i at the end of a word is usually spelled by the letter y.

Sub-test II- Proof Reading-- This test is designed to determine if children have a degree of visual perception which enables them to retain accurate impressions of words, and to discover if children can spell the commonly used words which are frequently misspelled. The words used in this test were selected from various lists of spelling demons or words frequently misspelled.

Sub-test III - Word Meaning-- Test on vocabulary consists of 30 lists of 6 words each. Each list contains several related words with unrelated words interspersed.

Sub-test IV- Handwriting-- The handwriting test consists of an original paragraph. The children are instructed to copy as much as they can in their usual quality and at their customary speed in two minutes. The words were taken from the Gates' list and were not above grade 6 difficulty.

Sub-test V- Visual Discrimination.-- The test consists of three parts. Part I and II on the pupil's test paper each contain fifteen rows of five words each. These words are from the Thorndike list and are from the tenth thousand or above in frequency. In Part I the teacher flashes a word for one second and the child is to find the word in a certain row and draw a line through it. In Part II the children wait for instructions to write. The wait is for five seconds. To determine the ability to retain visual impression of the word. In Part III after the five seconds have elapsed the child is directed to write the word from memory.

Sub-test VI- Auditory Discrimination.-- This test attempts to determine ability in auditory discrimination without any help from the visual. Part I is words from Gates and Buckingham-Dolch lists. Part II words are from the Thorndike list.

Samples of all the above listed tests are included.

Validity of tests:

1. Each test item was determined in relation to the best available studies in each phase of spelling.
2. Each word was checked or rechecked against several word lists for grade level and comprehension.

3. The preliminary and final testing showed that the form of the tests was suitable and convenient to administer.

Conclusions:

1. The test as a whole is valid and reliable.
2. The various sub-tests show a fair correlation except in handwriting, with spelling achievement, which shows that the test has diagnostic value.
3. The cause of poor spelling achievement may be disclosed through the results of the sub-tests.
4. The two forms provide for diagnostic and remedial procedures.
5. Correlation between handwriting and spelling achievement is positive but low, so low that it has doubtful value as a means of diagnosing spelling difficulties.

Nett, Helen Sullivan. The Construction and Evaluation of a Reading Vocabulary Test for Grades V, VI and VII. Ed.M. 1950. 59p.

Problem: The construction and evaluation of a sensitive vocabulary test involving varieties of word meanings in a shorter space and time.

Procedure: 1. A list of words was selected from several basal readers. These were checked with the Thorndike-Lorge list to determine grade frequency and range of recurrence. A list of synonyms were derived from Roget's Thesaurus.

2. The resulting list of 440 was given in January to 300 pupils, 100 pupils from Grades V, VI and VII. The pupils were from three public schools in a suburb of Boston.

3. Mental ages of the pupils were obtained from the Otis Quick Scoring Test given by the "Director of Pupil Placement." The results were as follows:

	Mental Ages		
	<u>Grade V</u>	<u>Grade VI</u>	<u>Grade VII</u>
Highest	13-5	14-6	15-8
Medium	10-5	11-2	12-3
Lowest	8-0	8-7	9-1

4. After obtaining the scores an item analysis for Grades V and VII was made in order to get the number of correct responses on each word for each grade. Tables were set up showing the percentage of correct responses on each word for each grade. Tables were set up showing (1) a summary of the Most

and Least Sensitive Items, (2) Words Too Easy, (3) Words Too Hard, (4) Weak Words, (5) Ambiguous Words, (6) Strong, Best Fitted Words.

5. Tables were also set up showing the correlation between the scores of the test and those of the reading achievement.

Conclusions: 1. No pupil obtained a perfect score. This indicates that an upper limit was successfully established. The scores are as follows: (highest possible score 200)

	<u>Grade VII</u>	<u>Grade VI</u>	<u>Grade V</u>
Highest	185	180	138
Mid-Score	148	109	87
Lowest	47	57	53

2. The correlations between the scores of the test and the reading achievement are as follows:

Grade Equivalent Scores Reading Metropolitan
Achievement Test, December 1949

	<u>Grade VII</u>	<u>Grade VI</u>	<u>Grade V</u>
Highest	8.9	7.3	6.7
Mid-Score	7.3	6.0	5.2
Lowest	4.6	5.4	4.6

3. An analysis of the scores indicates that this test would be more adaptable to Grades VI, VII and VIII. Its difficulty for Grade V may be due to failures in word recognition, confusion from the long form of the lists and possibilities of being misled by similar phonetic sounds.

O'Sheasy, Edward Andrew. A Study of the Predictive Values of Certain Pre-Reading Tests Employed at the First Grade Level.

Ed.M. 1951. 76p.

Problem: To determine the relationship of the Gates Reading Readiness Test, the Pintner-Cunningham Primary Test and teacher-judgement to reading achievement.

Procedure: 1. The Pintner-Cunningham and the Gates Reading Readiness tests were given in Sept., 1949 to 209 first grade pupils in six elementary schools in a suburban area of a South Atlantic port city.

2. The Gates Primary Reading Test was given during the month of May, 1950.

3. In addition to these tests given the following information was provided for each child.

- a. Chronological Age
- b. Intelligence Quotient
- c. Mental Age
- d. Teachers' Estimate of Probable Success

4. Frequency distribution tables were set up and means and standard deviations of the means were computed for the factors mentioned above.

5. The predictive values of the tests used were determined by correlating scores on the Gates Reading Readiness Test and the Pintner-Cunningham Primary Test with the Gates Primary Reading Test. Multiple correlation was used to determine predictive powers of the Gates Reading Readiness Test when used

with mental age.

6. The importance of the different levels of intelligence differences in reading readiness were determined by a critical ratio method. The same method was used also in determining sex difference.

7. Teacher predictions were correlated with achievement scores to see how valid their conclusions were.

Conclusions: 1. The coefficient of correlation between the mean scores for the Gates Reading Readiness Test and the Gates Primary Reading Test was .55-.047. Correlation is not sufficiently significant to warrant general use in prediction of individual progress. A study of the scores of the two tests show that the readiness test scores, unless they approached the extremes of distribution, were not accurate indicators of future progress in reading. The Gates Readiness Test did predict success. Of 16 children scoring 80 or higher only two failed to make high corresponding scores on the achievement test. The test did not reliably predict success in the lower quartile, as the 21 children scoring less than 20 on readiness test, 7 of them were successful in the achievement test.

2. It is evident that mental age is a valid indicator of future success, but a questionable indicator of probable failure. The correlation between the mean mental age and the mean achievement score was .47-.015. It was found that a high mental age was followed by successful reading achievement as the 16 children with a mental age of above 99 months, all had adequate reading ability on Gates Primary Reading test. However, 41 of the 68 children whose mental ages ranged between 50 and 75 months,

achieved a satisfactory level in reading. This study questions the ready acceptance of a mental age of seven years or six years and six months as a prerequisite for successful reading achievement.

3. No appreciable gain is realized by combining scores of Mental Age and Reading Readiness for providing prognostic aid. The readiness test alone seems to provide adequate information. The multiple correlation between mental age, reading readiness, and reading achievement was found to be .58. This was higher than the correlation between mental age and reading achievement (.47-.015), but no higher than the correlation between reading readiness and reading achievement (.55-.047).

4. Teacher judgement is a highly reliable indicator of future achievement in reading. Of the methods of predicting reading achievement used, teacher prediction correlated most highly. The coefficient of correlation was .65-.040.

5. It was found that true differences did exist in favor of the girls in all cases, except the Gates Primary Reading Test by method of critical ratio. On the Gates Reading Readiness Test the Critical Ratio was 3.02 in favor of the girls. On the Gates Primary Test it was .52- which does not prove that girls are superior to boys.

6. By means of the method of critical ratio, it was found that true mean differences exist between high, low, and average intelligence groups. The comparison between high group and average group resulted in a critical ratio of 7.88 in favor of high group. The comparison of scores for high group and low group resulted in a critical ratio of 9.92 in favor of high group.

The comparison of scores in favor of the average group and low group resulted in a critical ratio of 3.54 in favor of the average group.

7. It is concluded that the tests used in this study help the teacher, but do not surpass her judgement in the prediction of first grade reading achievement.

Sherman, Dorothy E. The Construction and Analysis of a Spelling Readiness Test. Ed. W. 1950. 52p.

Problem: To construct and analyze a spelling readiness test, which might determine if a child is ready to begin formal spelling.

Materials and Procedure:

The writer wished to test quality of handwriting and skill in auditory discrimination and visual memory. The program was set up using the following tests:

Sub-test I - Letter Forms

Sub-test II - Initial Consonants

Sub-test III - Final Consonants

Sub-test IV - Initial Blends

Sub-test V - Final Blends

Sub-test VI - Phonograms

Sub-test VII - Visual Memory

The key words for the sub-tests were chosen from Durrell's Vocabulary List for Primary Grades which was selected from the Faucett-Maki list and checked against the Fitzgerald and the International Kindergarten list.

Sub-test I - Letter Forms - tests the quality of handwriting. It consists of responding in writing to the dictating of the letter names.

Sub-test II - Initial Consonants - tests the ability to write the letters for 18 initial consonant sound. All consonants were included, except q, z, and x.

Sub-test III - Final Consonants - tests the ability to write the letters for 13 final consonant sounds.

Sub - test IV - Beginning Blends - consists of writing the first two letters of key words. The blends were carefully selected as pertinent to this age.

Sub- test V - Final Blends - tests the ability of writing two letters for the final sounds of key words.

Sub - test VI - Phonograms - consists of writing words that are spelled phonetically and contain the common vowel sounds. These were carefully checked against several lists.

Sub - test VII - Visual Memory - tests the ability to reproduce words shown by the flash-card method.

Following a preliminary try-out and the resulting revisions, the final form was administered to 141 pupils of grade one in five classes in four cities and suburban towns.

The data was analyzed to find:

1. The difficulty of the test items on the total test.
2. The difficulty of the sections within the test.
3. Sex differences within the same areas.

Conclusions:

1. The item analysis of the 97 test items indicated that 86 items had a C.R. of 3.000 or above and are valid test items.
2. There was not significant difference between the boys and girls on any section of the test.

Simon, Margaret. The Construction and Evaluation of a Reading Vocabulary Test for the Middle Grades. Ed.M. 1951. 51p.

Problem: To construct a more sensitive reading vocabulary test for the middle grades and to evaluate its results.

Procedure: 1. The test constructed used both a multiple response technique and a matching technique. Instead of measuring a single meaning of a word the test tried to measure varied meanings and shades of meanings, plus synonyms.

2. The words used in the study were chosen from several basic fifth grade readers and supplements by words from Thorndike's list which had a 60 or more count in the J column. These words were then grouped according to Roget's categories. Both the words and categories were checked for frequency of use. The final list consisted of 420 words.

3. In addition to the test constructed the Otis Quick Scoring Mental Ability Test was also given. The tests were given to 200 pupils in a public school outside of Boston, 67 fourth graders, 75 fifth and 58 sixth.

4. After the tests were scored an item analysis was made and when all items had been tabulated, the numerical figures were converted to per cents because the numbers of cases in each grade differed. Frequency distribution tables were set up to show the range of scores on the two parts of the test in the three grade levels tested. Mean, Standard Error of the Mean, Standard Error and the Critical Ratio were computed for both parts of the test. A scatter diagram was made to show a comparison of the

Otis and Vocabulary Test.

Conclusions: 1. The charts and tables show that the test was able to help in discriminating between the poor and the able readers. The Medians for Part I show Grade IV - 63.90, Grade V - 73.55, Grade VI - 80.75. Grades V and VI had a critical ratio of 3.443 and Grade V and VI - .879. Part II did not show up as well as the first part of the test in any of the data, although it also shows a good difference between the grades. The medians are: Grade IV - 3.49, Grade V - 5.99, Grade VI - 9.10. The critical ratio for this part for Grades IV and V is 1.05 and for Grade V and VI is .82.

2. The test was found to have many weak items. In Part I, 18 groups out of 40 and 130 words out of 280 were strong and showed differences between the grades tested. In Part II, 22 items out of 40 had power of discrimination.

Spencer, Doris. The Effect of Reading Disability On Performance On The Terman-Merrill Revision Of The Stanford Binet Intelligence Tests. Ed. M. 1943. 54p.

Problem:

Primary

(a) To determine whether scores on 1937 Revision of the Stanford Binet Intelligence tests are lowered because of reading disability.

Procedure:

Method Used

- (a) Give Binet to 302 reading disability cases.
- (b) Compare difficulty of test items in each year level.
- (c) Compare this with Durrell's study of Binet's first scale.
- (d) Tables were made to show following information.
 - (1) Range of I Q's 63-163 50% between 90-109.
 - (2) Mental ages 6-20-50% between 7-10 mean M. A. 8.0.
 - (3) Chronological ages 37% 7-8 year levels.
 - (4) No. and % of passes at each level.
 - (5) Order of difficulty of items.
 - (6) Successes of girls
 - (7) Successes of boys
 - (8) Comparison of difficulty on 1937 scale and first scale.

Findings and Conclusions

- (1) Items dependent on reading ability were failed most often.

(a) Item 3 on year X, "Reading and Report", % of passes 26.2.

(b) Item 6 on Year VIII Level Memory For Sentences, % of passes 57.5.

(c) Items 1 and 6 on Year XII year level Vocabulary and Minkus Completion, % of passes 24.8 and 18.9.

(2) Order of difficulty did not follow pattern in Standardization.

(3) Boys achieved higher percent of passes than girls.

Varney, Elinor Caroline. An Experiment to Evaluate Some Techniques for Measuring the Knowledge of Word Meanings in the Fourth, Fifth, and Sixth Grades. Ed.M. 1945. 84p.

Problem: To evaluate some techniques for measuring the knowledge of word meanings in fourth, fifth, and sixth grades.

Materials: 1. 100 words from Durrell's Primary Reading Vocabulary.

2. Five tests of word meanings made by the author.
3. Webster's New International Dictionary.

Procedure: 1. Care was taken to get common words, each with several meanings and offering no reading difficulty.

2. Five tests of recognition type, using matching, multiple choice, and true-false techniques, were constructed, each testing various meanings of 20 words.

3. The meanings were checked with an unabridged dictionary omitting any meanings that were dialect, obsolete, colloquial, slang, rare, archaic, or of special occupational significance.

4. The tests were given to 155 children in grades four, five, and six in a large residential community by the regular classroom teacher and scored by the author. There was not a time limit on the test.

Major Findings and Conclusions: 1. Matching and multiple choice techniques are effective procedures for measuring the different levels of meaning of the same word.

2. The test having the highest per cent of valid items was a matching technique with multiple choice second.

3. The matching technique tested the most meanings for the space consumed with multiple choice second.

4. The true-false technique had the lowest item validity.

Waters, Mildred. A Partial Evaluation of an Experimental Pre-Reading Test. Ed.M. 1949. 139p.

Problem: To evaluate the Readiness Battery Pre-Reading Test by Constance M. McCullough and David H. Russell, an experimental test of Ginn and Company.

Procedure: 1. A description of the experimental test is as follows:

The test consists of 70 items which are divided into five subtests which measure a separate factor related to reading success.

Test 1. Vocabulary Readiness: Designed to meet vocabulary of the Ginn three pre-primers. The child marks the picture which he thinks is the word the examiner pronounces.

Test 2. Tactile Visual Readiness: Each item consists of a series of three word blocks, two of which are alike in shape. The child is asked to trace the configurations and mark the one that is different.

Test 3. Visual Readiness: The child is supposed to mark the word that is different in a series of three words.

Test 4. Auditory Readiness: Each item consists of a row of four pictures. The first picture is the key picture. The child selects from the other three pictures the picture which represents the word that rhymes with the word represented by the key picture.

Test 5. Comprehension Readiness: The test has two parts and each part has four items. The items are rows of pictures

each representing action. After the teacher finishes story child marks picture in each row in response to the teacher's directions.

2. The tests were given to 656 first grade pupils during the fifth and sixth week of the school year (Oct., 1948). The twenty-five cooperating classes were from schools from eleven different states. All classes were using the Ginn Basic Reading Program.

3. The mean chronological ages of the 656 cases was 75 months with a standard deviation of 4.5. One hundred and sixty-four intelligence quotients were supplied by the different school systems: 40 were obtained on Stanford Binet, 69 on Detroit First-Grade Intelligence Test, 55 were from Kuhlmann-Anderson Intelligence Test.

4. The tests were scored by the writer and frequency distribution tables were set up showing the central tendencies and variabilities of the scores.

5. The text-books from which the test was constructed were reviewed to see if the test had curricular or face validity. Research related to the specific readiness factors included in the test was investigated to justify the use of the factors in the test.

6. Intercorrelations were made between the sub-tests and correlations were made with other tests (Stanford-Binet, Kuhlmann-Anderson, Detroit First-Grade Intelligence, Metropolitan Readiness Test) to see if the test had statistical validity.

7. A comparison was made between the scores of the pupils who had kindergarten experience with those who had not.

Conclusions: 1. There was a skewed distribution in both the total test and sub-test scores. The range of scores in the total test was from 33-70, but approximately 73% of the scores fell between 60 and 70. The scores on the sub-tests show the same results. In Test 1, Vocabulary, 69.96% of the pupils received perfect scores. The range was from 8-16, the median was 16, the perfect score. In Test 2, Tactile Visual Readiness, the range was somewhat wider than Test 1. The range in score was from 1-15; 52.28% of the pupils received perfect scores. In Test 3, Visual Readiness, the median score of this test was 14; 44.51% of the total scores on this test was perfect showing better distribution than the previous tests. Test 4, Auditory Readiness, the range is from 3-15; 53.96% of the pupils received perfect scores. In Test 5, Comprehension Readiness, the range is from 8; the median score was 6. Only .05% received perfect scores.

2. The intercorrelations of the sub-tests indicate that the sub-tests measure different abilities, but this conclusion must be very tentative because of the lack of normality in test data. The intercorrelations of the sub-tests are as follows:

Test		<u>Voc.</u>	<u>Tactile</u>	<u>Visual</u>	<u>Auditory</u>	<u>Compr.</u>	<u>Total Test</u>
1			.13	.30	.17	.03	.18
"	2	.13		.32	.23	.08	.74
"	3	.30	.32		.34	.15	.77
"	4	.17	.23	.34		.15	.67
"	5	.03	.08	.15	.15		.45

3. The correlations between the sub-tests scores and the total test scores indicate that the total test scores were dependent chiefly on Tests 2-3-4. Test 1 did not contribute substantially to the discriminating data.

4. The low coefficients between the Pre-Reading Test and each of the three intelligence tests indicate that the experimental test measures factors other than those measured by the intelligence tests. They are as follows:

Stanford-Binet	.26 on 40 cases
Kuhlmann-Anderson	.31 on 55 cases
Detroit 1st. Gr.	.61 on 69 cases
Metropolitan Readiness	.70 on 40 cases

The results show that the Metropolitan Readiness measures and the experimental Pre-Reading Test measures somewhat similar factors.

5. The difference between the means of the scores of the kindergarten group (M. 63.4) and the non-kindergarten group (M. 61.8) and the critical ratio of 3.28 indicates that kindergarten experience may account for the higher scores of that group. The skewed nature of scores must be considered, however.

Wilking, Stephen Vincent. The Construction and Evaluation of a Measure of Reading Vocabulary. Ed.M. 1940. 49p.

Problem: To build a test which could be incorporated in a diagnostic reading test.

Procedure: 1. To solve the problem of testing more words in less space, the category method of testing was used. Roget's Thesaurus was consulted for the setting up of these categories. A fourth grade vocabulary constructed from the reading books by Durrell-Sullivan and the first and twentieth thousand words in the Teacher's Word Book was also used. From this study 24 categories were derived. Eighteen categories were chosen to be used in the Reading Vocabulary Test.

2. Two experimental forms consisting of 360 words each were built and were called Forms A and B.

3. These tests were given to a total of 356 students in four grades: 3-9-11 and 13. The score distribution on each grade level was analyzed and an item analysis was made.

4. From the analysis made, two new forms, Forms AA and AB were made and equated for difficulty and step-up interval each consisting of 90 words, as opposed to the 360 words in the experimental form.

5. Only Form A was evaluated. The reliability was tested by means of a multiple correlation between the scores made on three sheets of the test on a group of 80 students in the ninth grade. The validity was tested by constructing a criterion test to measure it by.

Conclusions: 1. The differences in the mean score on each grade for Form A show a good difference. The critical ratios are all statistically significant (between 7 and 9-8.2, between 9 and 11-4.2, between 11 and 13-10.6). This shows that the chances are 99.9 out of 100 that the obtained difference is a true one and not due to the operation of chance factors.

2. The experimental test Form A proved to be a fairly reliable instrument with a reliability in the vicinity of .70 for a small population on one grade level.

3. The test is a highly valid measure of reading vocabulary as compared against the constructed criterion. This criterion was found to have a reliability coefficient of .69. The validity coefficients on three groups, the ninth grade, college freshmen and the two groups combined, were .58, .73 and .76.

Young, Patience. A Comparison of the Vocabularies of Ten Standard Achievement Tests in Reading with the Primary Social Studies Books of the Curriculum Foundation Series.
Ed. M. 1951. 119p.

Problem: To compare these vocabularies to determine what percentage of words included in the tests were in the books and what percentage of the vocabulary of these books was tested.

Materials:

1. The following standard primary reading achievement tests: American School Achievement Test, Detroit Reading Test, Detroit Word Recognition Tests, DeVault Primary Reading Test, Gates Primary Reading Test, Lee-Clark Reading Tests, Los Angeles Primary Word Recognition Test, Metropolitan Achievement Test, Reilley's Primary Reading Test, and Unit Scales of Attainment.
2. The primary social studies books: Peter's Family, Hello, David, and Someday Soon.

Procedure:

1. The vocabulary list of new words at the back of each book was analyzed and the words arranged in alphabetical order, the number of words being put at the head of each list.
2. The words of all the tests were arranged alphabetically for each form and grade with the number of words heading each list.

3. The tests for the primer or first half of the first grade were compared with the Social Studies primer; for the first grade or last half of first grade with the Social Studies Book One; for the second grade or grades one and two, the three books were used.
4. Words common to both tests and texts were recorded on a work sheet. Percentages were determined as stated in the problem and these percentages arranged in tables for interpretation.

Major Findings and Conclusions:

1. There was a wide difference between the tests in per cent of words common to the tests and books.
2. The Detroit tests showed generally high percentages of words common to both. With the exception of the DeVault Primary Reading Test, Form A, which was second with 69.23%, the highest seven all were forms of Detroit tests. The lowest percent was found in the Unit Scales of Attainment, First Half, Grade One, Form A.
3. Percentages for the number of test words found in the books ranged from 70.31 to 27.76.
4. All of the indications pointed to the fact that these tests did not sample adequately the vocabulary of the books. Therefore, these ten standard tests did not seem suitable for use where the Curriculum Foundation Series of social studies books form an important part of the reading program.